THE EARTH

CAMBRIDGE
UNIVERSITY PRESS
LONDON · Fetter Lane
C F. CLAY, Manager

LONDON
H. K. Lewis & Co., Ltd.,
136, Gower Street, W.C. 1

NEW YORK
The Macmillan Co.

BOMBAY, CALCUTTA and
MADRAS
Macmillan and Co., Ltd.

TORONTO
The Macmillan Co. of
Canada, Ltd.

TOKYO
Maruzen-Kabushiki-Kaisha

THE EARTH

ITS ORIGIN, HISTORY AND PHYSICAL CONSTITUTION

By HAROLD JEFFREYS, M.A., D.Sc.
Fellow & Lecturer of St John's College, Cambridge

CAMBRIDGE: AT THE
UNIVERSITY PRESS
MCMXXIV

PRINTED IN GREAT BRITAIN

TO THE MEMORY OF THE LATE
SIR GEORGE HOWARD DARWIN
THE FOUNDER OF MODERN
COSMOGONY AND GEOPHYSICS

PREFACE

During the years 1920–23 I delivered three times in this college a course of eight lectures on the Physics of the Earth's Interior. The course was intended to give an outline of present knowledge concerning what may be called the major problems of geology, namely the physical constitution of the earth, the causes of mountain formation, and the nature of isostasy. It was, however, impossible to give in the lectures full accounts of the arguments employed, partly because the course was too short, and partly because the mathematical knowledge of the listeners was extremely varied. Accordingly this book has been written; the argument has been made as connected as appeared possible, and various geophysical topics that could not be discussed in the lectures, such as the variation of latitude, have been introduced. The aim has been to discuss the theories of the main problems of geophysics, and to exhibit their interrelations. Several large branches have, however, been almost entirely omitted; terrestrial magnetism, atmospheric electricity, tidal theory (apart from tidal friction) and meteorology have received little or no attention, because to give anything approaching an adequate account of any of them would have required a longer discussion than their connexion with the original topics of the book seemed to warrant.

I have attempted to describe the present position of the subject, rather than its history. For this reason several pieces of work of capital importance in the development of geophysics have escaped mention. Lord Kelvin's estimate of the rigidity of the earth from its tidal yielding is a case in point; it has not been discussed because more detailed and definite information can now be derived from seismology. Sir G. H. Darwin's pioneer work on tidal friction, again, has been only outlined, because it was mainly a discussion of bodily tides in a homogeneous earth, which now appear to be comparatively unimportant in influencing the evolution of the earth and moon. Nevertheless if the contents of the second volume of Darwin's collected papers had not been published, it is improbable that Chapters III and XIV of this book would have been written.

Quantitative comparison of theory with fact has always been the main object of the book, and practically all the theories advocated have survived the test of quantitative application to several phenomena. Accurate theories have been given where they seemed necessary for this purpose; but where an estimate of an order of magnitude was all that was required, and could be obtained by rough methods, such methods have always been used. I have been encouraged in the latter course by several facts. First, though the method of orders of magnitude is not convincing to the pure

mathematician, it is a matter of experience that when a problem discussed by this method is afterwards solved by more formal methods, the answer is found to be of the correct order of magnitude, which is all that the method claims: it could be vitiated only by a fortuitous numerical coincidence. Second, though in some cases formally accurate solutions of related problems exist, or could be obtained, the problems actually so soluble differ so much from those that actually arise in geophysics that, in their actual application, they could at best be correct only as regards order of magnitude. Thus they are not inherently any more informative than the rougher methods. Third, a direct proof that a particular hypothesis will account for particular data is not very strong confirmation of the hypothesis when both the data and the consequences of the hypothesis are known only vaguely; but if it is shown that the results of the hypothesis agree with the facts as regards order of magnitude, while the results of denying it are in definite disagreement, the confirmation of the hypothesis will be almost as strong as if a close agreement had been obtained. The method of exhaustion of alternatives is specially useful in geophysics, because incorrect geophysical hypotheses usually fail by extremely large margins.

Two criticisms are certain to be made by geologists, and therefore I venture to attempt to meet them in advance. The first is that the book contains a great deal of matter not of a geophysical character. I thought at first of avoiding this objection by dividing the book into two parts, one cosmogonical and one geophysical; but I found such a course impossible, since the two were too closely interwoven, each depending in part on the results of the other. In a work mainly theoretical rather than descriptive in character it therefore seemed best to develop the implications of a hypothesis wherever they might lead to results capable of empirical test, rather than to confine my attention to one particular planet. If a theory is satisfactory, the more it is shown to explain the more reliable it is; and if it is unsound, it will be unsound whether the fact inconsistent with it happens to relate to the earth or to the satellites of Uranus.

The other objection I anticipate is that the book is too mathematical for geological readers. The answer is simple: the results aimed at are quantitative, and there is no way of obtaining quantitative results without mathematics. I have tried to keep the mathematics as elementary as possible; but some problems could not be handled by simple mathematics, and I had no alternative except to give all that was necessary. If the geologist cannot follow a part of the book, I hope he will omit it and go on to the next non-mathematical passage, trusting that someone else will point out any intervening mistakes (and the mathematically trained readers, with few exceptions, will do just the same). He will then know that at any rate some people will be able to follow the argument all through, and he will see just where he fails to follow it himself; whereas

a so-called non-mathematical "exposition" would only bewilder the mathematical physicist, while making it impossible for the geologist to see how much is hypothesis and how much is merely the systematic investigation of the consequences of hypotheses already made and data already found. In short, if geophysics requires mathematics for its treatment, it is the earth that is responsible, not the geophysicist.

The paragraphs are numbered according to the decimal system; of any two paragraphs, that with the smaller number comes earlier in the book. The integral part of a paragraph number is the number of the chapter. Equations have as a rule been numbered consecutively through each paragraph; but in some cases they have been numbered consecutively through several closely related paragraphs. In cross-references, where reference is made to another equation in the same paragraph, only the number of the equation is given; but where the reference is to a different paragraph, the numbers of that paragraph and of the equation are given; *e.g.* 14·61 (3).

I wish to express my thanks to the staff of the Cambridge University Press for their courtesy during the publication of this book; to Mr R. Stoneley, who has read the whole in proof and checked a great deal of the mathematical work, suggesting many improvements in the process; to Dr J. H. Jeans, who verified Chapter VII; to Dr A. A. Griffith, who gave me much of the information incorporated in Chapter IX, though he does not wholly approve of my terminology; to Dr Arthur Holmes, whose influence on my geophysical thought has been none the less important because I experienced it before beginning this book; and to Sir Ernest Rutherford, Prof. Eddington, Prof. Shapley, and Dr Wrinch, who have read various portions in typescript and suggested improvements.

HAROLD JEFFREYS

ST JOHN'S COLLEGE
CAMBRIDGE
1924 *March* 6

CONTENTS

INTRODUCTION

"I am doing the best I can at my age."
G. B. SHAW, Preface to *Back to Methuselah.*

THE special difficulty of the problem of the physics of the earth's interior arises from the extremely restricted range, both of depth and of time, for which any direct evidence is available. The deepest boring yet made is one of 2·5 km., and there are few over a kilometre in depth; mines are still shallower; and apart from borings and mines the only accessible parts of the crust are those exposed in quarries and cliffs, and the land surface itself. Samples of material from greater depths are continually being brought up to the surface, but the possibility of utilizing any information they may give us is seriously hampered by the fact that we know only vaguely from what depths they have come and to what extent their temperature and even their chemical constitution have been altered in their ascent. In addition, all observations of earthquakes and all observations of topography and gravity have to be made at the surface or near it. Thus the problem of the physics of the earth's interior is to make physical inferences over a range of depth of over 6000 km. from data determined only for a range of 2 km. at the outside.

Extrapolation over such a range can be carried out only if we have some extremely reliable scientific laws, well verified in the laboratory and the observatory, to guide us in our investigation. Fortunately this is the case. Considerable headway can be made by means of such laws as those of gravitation, heat conduction, radioactive disintegration, elasticity, and fluid friction, all of which are well established. It is naturally impossible to proceed far in such a subject without introducing, besides known laws and empirical data, some *ad hoc* assumptions. The number of these has to be restricted as severely as possible; the justification of any hypothesis lies in its ability to coordinate otherwise unrelated facts, and hypotheses that do not explain more than they assume should be omitted.

In many questions relating to the interior of the earth some knowledge of its thermal state is indispensable. There is no direct method of arriving at this knowledge; but we may make use of the laws of heat conduction to arrive at it indirectly. Given the distribution of temperature in a solid at any instant, and the thermal conductivity, it is possible to find the distribution at any subsequent instant, and therefore if we know the thermal conditions inside the earth at some past date it should be possible, given sufficient mathematical labour, to find those of the present time. We have not such knowledge for any historical date, but we can find it for the time when the earth was first formed, by making use of the modern

theories of the origin of the solar system. It is therefore necessary to give first of all an account of the evidence in favour of these theories; but in so doing we must make use of information relating to the other members of the solar system besides the earth. It is found that the theory finally selected as the most satisfactory gives fairly definite information about the initial thermal state of the earth, which we can use in determining its subsequent history.

The thermal history of the earth is influenced not only by the initial distribution of temperature, but also by radioactive substances in the crust, which are responsible at the present time for the greater part of the heat that is continually being conducted up to the surface and finally radiated away. It is therefore necessary to make a great deal of use of the information supplied by geologists concerning radioactivity in rocks, especially in relation to the age of the older rocks and the present thermal output. The thermal history in its turn supplies information about the elastic properties of the interior of the earth and about the deformations set up by contraction and other disturbing factors. In this way it is found possible to give a coherent account of the dynamics of isostasy, in relation to the support of both continents and mountains.

We wish, then, to arrive at a decision concerning the primitive state of the earth in order to utilize it in an account of the subsequent history and of the present physical condition. The chief questions to be dealt with are first, whether the earth was ever mostly fluid, or whether it has always been solid throughout, and cool near the surface; and second, if it be granted that it was ever fluid, to form some conception of the manner of solidification and the physical conditions just after solidification. A direct answer to either of these questions requires a discussion of the origin of the earth, concerning which numerous conflicting hypotheses have been propounded. A stage in the development of cosmogony has, however, now been reached in which both points can be decided with a high degree of probability.

Practically all hypotheses regarding the origin of the solar system agree in supposing the sun and the planets to have formerly formed a single mass; they differ in the processes supposed to have led to the separation of the mass into many parts and in the subsequent evolution of the planets. Most of the hypotheses may be divided into two main groups, which may be called the Rotational and Tidal types. In the rotational type the rupture is supposed to have arisen through the speed of rotation increasing, as the mass condensed, to such an extent that gravity ceased to be adequate to hold it together. In the discussion of this type the leading investigators have been Laplace, Babinet, Roche, Moulton and Jeans. In the tidal type a star larger than the sun is supposed to have come close to the sun and practically torn it asunder by its tidal action. A hypothesis resembling this was advanced by Buffon, but fell out of

favour after the criticisms of Laplace. Its modern development was due in the first place to Chamberlin and Moulton; but several features of their formulation have been found unsatisfactory by the present writer, who has therefore attempted to reconstruct the theory entirely from the time of the rupture. At the same time Jeans has carried out a comprehensive and general dynamical discussion of the possible methods of break-up of fluid masses, and has developed in it an account of the processes concerned in the formation of a system by tidal action. The theory he has reached bears a close resemblance to mine, the agreement being especially striking since the two methods of attack are quite different. The theory adopted in this work will be of the tidal type, and based on a comparison of those already advanced by Jeans and myself. It is not, however, possible to give an adequate statement of the reasons for adopting this particular theory without first giving some account of the advantages and drawbacks of the earlier theories of the origin of the solar system. Descriptions will therefore be given of the two chief theories that preceded the present one, namely the Nebular Hypothesis of Laplace and Roche, and the Planetesimal Hypothesis of Chamberlin and Moulton. It has, however, been found more convenient to treat the latter in an Appendix than to place it in its proper historical position, which would have been immediately after the Laplacian theory.

The tidal theory of the origin of the solar system leads directly to the inference that the earth was formerly fluid, and hence to an account of its method of solidification and of the distribution of temperature immediately afterwards. The age of the earth and the amount of radioactive matter near the surface are found from modern information about radioactivity in rocks, and these data together are used to infer the thermal history of the earth since solidification, and incidentally the distribution of temperature at the present time. It is then shown that the observed phenomena of mountain building and isostasy are substantially what might be expected from the theory of the cooling of the earth. The evidence supplied by seismology, the figures of the earth and moon, tidal friction, and the variation of latitude is discussed in subsequent chapters, and its relation to the thermal state of the earth and the evidence of geodesy is examined.

Some topics of geophysical interest, though outside the main argument of the book, are discussed in appendices. These are the Planetesimal Theory, Jeans's Theory (points not discussed in Chapter II), the Hypothesis of the Indefinite Deformability of the Earth under Small Stresses, Theories of Glaciation, and Empirical Periodicities.

The main purpose of the book is to show how far the various modes of attack on the problems of geophysics have led to harmonious results. Accordingly information from many widely separated fields has had to be collected. Probably the range of the subjects discussed is so

wide that nobody could possess a specialist's knowledge of all; I am quite certain that neither I nor anybody else possesses such knowledge at present. Imperfections are therefore unavoidable. Nevertheless I think that geophysics has suffered in the past from the lack of any systematic attempt to coordinate its various branches, and that many wild theories might have remained unpublished or, if published, have received only their fair share of attention, had it been possible to see at once how far they conflicted with the data of related branches of the subject. If therefore this book fails, as it must, to instruct the specialist in the data of his own subject, but sometimes enables him to check a hypothesis by means not only of his own data, but by some of those of related subjects, it will have served its purpose. Geophysics is no longer a field for uncontrolled speculation; it is a single science whose data are harmoniously coordinated by a definite physical theory, and any theory that contradicts this one must be shown to coordinate all these data equally satisfactorily before it can be accepted.

CHAPTER I

The Nebular Hypothesis of Laplace

"Though the mills of God grind slowly, yet they
grind exceeding small." LONGFELLOW.

1·1. *Laplace's Account of the Nebular Hypothesis.* The Nebular Hypothesis is usually attributed to Laplace, but the account of it given by him was exceedingly vague, no quantitative discussion whatever being included. It occurred in a semi-popular work, entitled *Exposition du système du monde,* the second edition of which was published in the seventh year of the French Revolution. The following account is a translation of that given by Laplace in the sixth chapter of the fifth book of this work:

Let us now pause to consider the arrangement of the solar system and its relations to the stars. The immense globe of the sun, the focus of the planetary movements, turns on its axis in twenty-five and a half days. Its surface is covered by an ocean of luminous matter whose active effervescences form variable spots, often very numerous, and sometimes larger than the earth. Above this ocean, a vast atmosphere rises; it is beyond it that the planets, with their satellites, move in almost circular orbits, in planes slightly inclined to the solar equator. Countless comets, after approaching the sun, depart to distances that prove that its empire extends much further than the known limits of the planetary system. Not only does the sun attract all these globes, leading them to move around itself, but also it sheds on them its light and its heat. Its beneficent action produces the growth of animals and plants on the earth, and analogy leads us to think that it produces similar effects on the planets; for it is not natural to think that the matter whose fertility we see developed in so many ways is sterile on so great a planet as Jupiter, which, like the terrestrial globe, has its days, its nights, and its years, and on which changes implying very vigorous activity are indicated by observation. Man, made for the temperature he enjoys on the earth, could not, to all appearance, live on the other planets; but must there not be endless organisms adapted to the various temperatures of the globes of this system? If the difference of elements and climates alone can make such variety in terrestrial productions, how much more must those of the various planets and their satellites differ? The most lively imagination can form no idea, but their existence is very probable.

Although the elements of the system of the planets are arbitrary, there are nevertheless very remarkable relations between them that may enlighten us with regard to their origin. It is astonishing to see that all the planets move around the sun from west to east, and almost in the same plane; all the satellites move around their primaries in the same direction and almost in the same plane as the planets; and finally, the sun, the planets and the satellites whose rotations have been observed turn on their axes in the same direction, and nearly in the plane of their revolutions in their orbits.

So extraordinary a phenomenon is not the result of chance; it indicates that a general cause has determined all these movements. To find approximately the probability with which this cause is indicated, we observe that the planetary

system as we know it to-day is composed of seven planets and eighteen satellites; the rotations of the sun, five planets, the moon, the satellites of Jupiter, Saturn's ring and its most remote satellite, have also been observed. These movements together form a set of thirty-eight movements in the same direction, at least when one refers them to the plane of the solar equator, to which it seems natural to compare them. If we consider the plane of any direct movement, lying at first upon the plane of the sun's equator, and then becoming inclined to this last plane, and traversing all the degrees of inclination from zero to two right angles, it is clear that the movement will be direct for all inclinations less than a right angle and retrograde for greater inclinations. Thus all motions, direct and retrograde, can be represented by a change in inclination alone. The solar system, looked at from this point of view, offers then thirty-seven movements whose planes are inclined to that of the solar equator by at most a right angle. Supposing that their inclinations are due to chance, they could have extended to two right angles; and the probability that at least one of them would have exceeded one right angle would be $1 - \frac{1}{2^{37}}$ or $\frac{137438953471}{137438953472}$. It is then extremely probable that the direction of the planetary movements is not at all the effect of chance, and this becomes still more probable when we realize that the inclinations of the majority of these movements to the sun's equator are very small and much less than a right angle.

Another equally remarkable characteristic of the solar system is the smallness of the eccentricities of the orbits of the planets and their satellites, whereas those of the comets are much elongated. No intermediate stages between large and small eccentricities occur in the system. We are then again compelled to recognize the effect of a regular cause. Chance alone would not have given an almost circular form to the orbits of all the planets, and the cause that determined the movements of these bodies must therefore have rendered them nearly circular. This cause again must have influenced the great eccentricity of the orbits of the comets, and, what is very extraordinary, without having affected the directions of their movements; for when we consider the orbits of retrograde comets as inclined at more than a right angle to the plane of the ecliptic, we find that the mean inclination of the orbits of all observed comets is almost a right angle, as it would be if these bodies had been projected at random.

Thus, returning to the cause of the primitive movements of the planetary system, we have the five following facts:

1. The movements of the planets are in the same direction, and almost in the same plane.

2. The movements of the satellites are in the same sense as those of the planets.

3. The rotations of these different bodies and the sun are in the same direction as their revolutions and in only slightly different planes.

4. The eccentricities of the orbits of the planets and their satellites are small.

5. Finally, the eccentricities of the orbits of comets are great, although their inclinations can be attributed to chance...*.

Whatever may be its nature, since it has produced or directed the movements of their satellites, this cause must have been common to all these bodies;

* Here Laplace interpolates a short discussion of Buffon's hypothesis, which will be dealt with in connection with the tidal theories.

and seeing the enormous distance that separates them, it can have been only a fluid of immense extent. To have given them an almost circular motion around the sun in a single direction, the fluid must have surrounded the sun like an atmosphere. Consideration of the planetary movements leads us then to think that in consequence of an excessive heat the atmosphere of the sun formerly extended beyond the orbits of all the planets, and that it has since withdrawn to its actual limits. This might have taken place through causes similar to that which produced the brilliant outburst in 1572, lasting several months, of the famous star in the constellation Cassiopeia.

The great eccentricity of the orbits of comets leads to the same result. It evidently implies the disappearance of a large number of less eccentric orbits; which supposes an atmosphere around the sun extending beyond the perihelia of the observable comets, which, by destroying the movements of those that traversed it during its great extent, has reunited them to the sun. Thus we see that only the comets that were beyond in that interval should exist at the present time; and as we can observe only those which approach sufficiently near to the sun at perihelion, their orbits must be very eccentric. But we see at the same time that their inclinations must offer the same irregularities as if they had been projected at random, since the solar atmosphere has in no way affected their movements. Thus the long periods of revolution of comets, the great eccentricities of their orbits, and the variety of their inclinations, are very naturally explained by means of this atmosphere.

But how has it determined the revolutions and rotations of the planets? If these bodies had penetrated into the fluid, its resistance would have made them fall into the sun; it may then be conjectured that they were formed at the successive limits of this atmosphere, by the condensation of the zones that it must have abandoned in the plane of its equator while cooling and condensing towards the surface of that luminary, as has been seen in the preceding book. We may conjecture again that the satellites have been formed in a similar manner by the atmospheres of the planets. The five facts asserted above follow naturally from these hypotheses, to which the rings of Saturn add a new degree of plausibility. Finally, if in the zones abandoned in succession by the solar atmosphere, there were some molecules too volatile to unite with each other or to heavenly bodies, they must, while continuing to revolve around the sun, present to us all the appearances of the zodiacal light, without offering any noticeable resistance to the movements of the planets.

Whatever may become of this genesis of the planetary system, which I present with the lack of confidence that everything that is in no respect a result of observation or calculation must inspire, it is certain that its elements are so regulated that it must enjoy the greatest stability, so long as outside causes do not disturb it.

1·2. *The Dynamics of the Nebular Hypothesis.* The purely qualitative discussion given by Laplace is from its nature necessarily extremely incomplete, and in the absence of quantitative treatment it was, as he says, impossible to feel much confidence in the hypothesis. He never carried out such treatment, and for over sixty years the question of the mode of formation of the solar system remained as he had left it. Before proceeding to the later investigations, however, it is desirable to dwell further on Laplace's account, and to point out the places where his

development is incomplete and the modifications produced by modern observation in the empirical data he used.

The solar atmosphere whence the planets are supposed to have been formed was originally a highly diffuse nebula; this was apparently considered to have been approximately symmetrical about an axis nearly perpendicular to the plane of the ecliptic, since it is described as extending beyond the orbits of all the planets and as bearing some resemblance to Saturn's ring. The motion of the nebula is only stated to have been one of rotation; it is not made clear whether the periods of revolution of all parts of the nebula were to be the same, or whether, as might seem possible, the inner parts would revolve more quickly than the outer ones. The distribution of density is not definitely specified; but the description of the nebula as surrounding the sun like an atmosphere suggests that the sun was already in a fairly advanced state of condensation, and that the total mass of the nebula was probably less than that of the sun.

The nebula condensed slowly, and its rotation became faster and faster, in accordance with the dynamical principle of the conservation of angular momentum. Laplace assumes that his system is unaffected by outside disturbances. In such a system suppose the mass of some particle to be m, and consider the motion of the foot of the perpendicular from that particle to some fixed plane. Suppose this point to be moving with velocity v, and that the perpendicular from a fixed point in the plane to the tangent to the path is of length p. Then the principle asserts that if the values of mpv for all particles of the system are added up, the sum is constant for all time, no matter what changes occur within the system. This sum is called the angular momentum of the system about the given axis. Let us now apply this principle to the nebula. If r denotes the length of the perpendicular from a particle on to the axis of rotation at time t, while the fixed point and plane are the centre and the equatorial plane respectively, the sum is Σmrv, where as usual Σ is used to denote 'the sum of the values of.' Now suppose that the contraction takes place in such a way that the values of r for all parts of the nebula are altered in the same ratio. If a particle was initially at distance r_0 we shall then have

$$r = r_0 f(t),$$

where $f(t)$ depends on t alone. Suppose that the same holds for the velocity, so that if the initial velocity of a particle was v_0, we have

$$v = v_0 g(t).$$

As the mass of any particle remains constant, it follows from the above principle that

$$\Sigma m r_0 v_0 f(t) g(t) = \Sigma m r_0 v_0.$$

But since r_0 and v_0 do not depend on the time, we must have

$$f(t) = 1/g(t).$$

Now if the equatorial radius of the nebula is a, it is clear that a is pro-

portional to $f(t)$, and that the velocity at the outer boundary is proportional to $g(t)$ and therefore to $1/a$.

The outer portions of the nebula at any time would be describing circular paths about the axis, under the gravitational attraction towards the interior and the fluid pressure acting outwards. The acceleration towards the axis required to maintain a circular path is v^2/r and is therefore proportional to a^{-3}. Since the action of fluid pressure is always outwards, rupture will take place if this acceleration is greater than that produced by gravity, namely a constant multiple of M/a^2, where M is the mass of the whole system. Hence as the mass contracts the acceleration necessary to hold it together increases like a^{-3}, while that available for the purpose only increases like a^{-2}. Therefore whatever may have been the initial velocity of rotation of the mass, it is only a question of waiting till it has contracted sufficiently for gravity to become inadequate to hold the outermost portions in contact with the remainder. When this is so, they will be left behind. Rupture must take place at this stage unless it has previously taken place in some other way.

The foregoing argument is of considerable generality, for it covers the extreme cases where the mass rotates like a rigid body (that is to say, all parts take the same time to revolve around the axis) and either is homogeneous or has nearly the whole of the matter concentrated into a central nucleus. An indefinite number of possible intermediate states are also covered, and also a still wider range of states where the motion is not like that of a rigid body.

Given some rotation to start with, the advance of condensation will necessarily tend to produce some kind of rupture. It may happen that the nebula will become liquid or solid before conditions suitable for rupture are reached; condensation will then cease and rupture will be postponed indefinitely. A certain minimum angular momentum is therefore necessary to produce ultimate rupture in any given mass. But supposing such an amount of angular momentum to be available, rupture is certain to occur before the whole mass has been absorbed into the central body. It may take place at an earlier stage than is indicated by the theory just given. If this is so, it happens while the gravitational acceleration available at the outer boundary is adequate to retain the matter there. Hence it takes place, not by crossing the outer boundary, but by internal condensation.

Laplace does not consider the possibility of the formation of planets by internal condensation, but only the separation of the outer portions around the equator. The whole system is perfectly symmetrical in these circumstances, and therefore the detached matter would form a ring. As condensation proceeded, other rings would be formed, and after its separation each ring would revolve independently about the central body with the period appropriate to its distance. Several objections have been

made against this part of the theory. It was suggested* that the result would not be a series of thick rings, nine in number, but a very large number of extremely thin ones, thus tending to account for minor planets and meteors rather than major planets. Other writers have argued that the absence of cohesion would make the separation quite continuous, whereas the formation of separate rings requires it to have been intermittent. These objections were partially answered by Roche†, in his detailed discussion of the Nebular Hypothesis, the irregularity being attributed to intermittent cooling of the central body; but his explanation seems very artificial.

Granting that separate rings could have been formed, it remains to show how a ring could have condensed into a single planet. On this point neither Laplace nor any of his followers offers any clear account. Such condensation appears to be impossible, in consequence of certain results obtained by J. Clerk Maxwell in his *Essay on the stability of the motion of Saturn's Rings*, published in 1859. He shows that a ring of particles surrounding a central body will be stable if the number of the particles does not exceed a limit depending on the mass of the ring, and unstable if the number is greater than this limit. In a fluid ring the number of particles is effectively infinite, and therefore the motion of such a ring is necessarily unstable, and the ring will break up. Now any disturbance of the motion of such a ring can be represented as composed of a number of waves travelling around it. Each wave will remain steady, die down, or increase independently of all the others. Maxwell shows that the shorter the wave length the more rapidly will a disturbance increase, and therefore the method of rupture is that the fluid ring first breaks up into a very large number of detached bodies; the disturbances of somewhat longer wave lengths then assert themselves and cause the masses to run into each other and combine until the number of separate masses is so reduced that the condition for stability is satisfied. Thus the rupture of a ring would lead to the formation of a ring of planets in nearly the same orbit and of comparable size, which is not a state now represented in the solar system. If Jupiter were divided into 49 fragments revolving about the sun, it would satisfy Maxwell's criterion of stability, and hence when this stage had been reached further aggregation would cease. Thus the formation of the solar system by the breakdown of rings is impossible.

Even at this stage the questionable assumptions of the hypothesis are not exhausted. If the planets were formed by this process, the fact that they all revolve in one direction is explained; so are the small inclinations and eccentricities of their orbits. But when we come to the

* Daniel Kirkwood, *M.N.R.A.S.* **29**, 1869, 96–102.

† 'Mémoire sur la figure des atmosphères des corps célestes,' *Mémoires de l'Acad. de Montpellier*, **2**, 1854, 399; 'Essai sur la constitution et l'origine du système solaire,' *loc. cit.* **8**, 1873, 235. H. Poincaré, *Leçons sur les hypothèses cosmogoniques*, 1913, 15–67.

satellites there are further difficulties. Each planet is supposed to have acquired, by some unspecified process, a rotation in the same direction as its revolution, and the condensation of the planets is supposed to have afterwards led to the formation of systems of satellites by the same mechanism as produced the planets in the first place. Thus the fact, pointed out by Laplace, that most satellites revolve in the same sense as their primaries, is explained. But unfortunately for the theory eight satellites now known revolve in the opposite direction, a fact which is quite inexplicable by this means alone. It has been suggested that the rotations of the planets were originally retrograde, but were reversed by solar tidal friction during the condensation, but it has been found that this is quite inadequate to account for the direct rotation of Jupiter and Saturn (see **13·73**).

Even in the case of the comets the theory is of doubtful value. The action of a resisting medium near perihelion would gradually reduce the mean distances of all comets that ever entered it, and therefore all surviving comets must have had perihelion distances greater than the present distances of Neptune. A further explanation of how these bodies came to be deflected so as to have small perihelion distances is therefore required, and is not provided by the theory.

1·3. *The Nebular Hypothesis with Internal Condensation.* Although the particular course of evolution of the solar nebula sketched out by Laplace has proved to be impossible, it does not follow that an extended nebula could not develop into a system very like the solar system by some different process. It is noticed on examining Laplace's discussion that the abandonment of perfect symmetry takes place at the very last moment conceivable, and that so long as we admit only symmetrical forms the only course possible is one in which the nebula sheds matter continuously around its equator, thus forming a new nebula around itself with a different distribution of velocity and probably of density. Now to produce the planetary system, symmetry must be abandoned at some stage, and it has been shown that if the abandonment is postponed till rings have been formed it is by that time too late. Emission from the boundary is not possible in any other way; the only alternative left is by internal condensation, a possibility that has already been suggested. Such a theory would require that some slight inequality of density should be formed within the nebula, and that on account of a type of instability the inequality should proceed to increase until the form of the nebula was very seriously altered and one or more planetary nuclei formed. We see at once that this accounts for several features of the system, if it is a possible mode of development. The direct revolution of the planets in slightly inclined and slightly eccentric orbits is explained. The rotation of a planet and the revolution of its satellites would be in the direction of rotation of the matter in the

neighbourhood of the primitive condensation, and would be direct if the condensation started within the inner portion, but perhaps reversed if within the matter already shed from around the equator. Thus it might be possible to explain why the most remote planets have retrograde satellites; but there appears to be still no way of explaining the existence of both direct and retrograde satellites of the same planet, as is true of both Jupiter and Saturn.

1·31. This hypothesis, however, fails on other grounds. The principle of the constancy of angular momentum, used at the very outset, imposes a severe restriction on the initial conditions of the mass, which turns out to be inconsistent with the condition that condensation shall be possible at all. If we use astronomical units of length, time, and mass, so that the earth's distance from the sun and the sun's mass have measure unity and the year has measure 2π, we can calculate the angular momentum of the whole system at present about an axis through the centre of the sun perpendicular to the ecliptic, and this cannot have changed since the earliest times. The data for the various planets are as follows:

Body	Mass $1\div$	Mean distance	Mean motion	Angular momentum
Sun	1	$2{\cdot}7\times10^{-3}$	14·4	$1{\cdot}05\times10^{-4}$
Mercury	9700 000	0·39	4·17	$6{\cdot}5\times10^{-8}$
Venus	402 000	0·72	1 63	$2{\cdot}1\times10^{-6}$
Earth	329 000	1·00	1·00	$3{\cdot}0\times10^{-6}$
Mars	3100 000	1·52	0 53	$4{\cdot}0\times10^{-7}$
Jupiter	1047	5·20	0 0844	$2{\cdot}18\times10^{-3}$
Saturn	3510	9·54	0·0367	$0{\cdot}95\times10^{-3}$
Uranus	22800	19·2	0·0119	$0{\cdot}19\times10^{-3}$
Neptune	19700	30·1	0·0061	$0{\cdot}28\times10^{-3}$

In this table the mean motion n is the mean angular velocity of the body as seen from the sun; it is also the reciprocal of the number of years in the period of revolution. The velocity is therefore rn, and the angular momentum is mr^2n. For the sun the mean distance given is k, the radius of gyration, defined as the distance from the axis of a particle with the whole mass of the sun which would have the proper angular momentum if it revolved in the sun's period of rotation. It is supposed that the density of the sun is uniform, so that

$$k^2 = \tfrac{2}{5}r_0^2 \qquad \text{.........(1)},$$

where r_0 is the radius of the sun. If the sun is condensed towards the centre this value will be reduced. The mean motion given for the sun is its angular velocity of rotation.

When we add up all the angular momenta, the total is found to be $3{\cdot}7\times10^{-3}$, of which Jupiter contributes more than half and the sun only about a thirtieth. The contributions of the inner planets, the satellites and the rotations of the great planets on their axes are all inappreciable. Now if the sun and all the planets were united into a homogeneous sphere or ellipsoid filling the orbit of Neptune and rotating in Neptune's actual

period, its angular momentum would be $\frac{2}{5}Mr_9{}^2n_9$, where M is the total mass of the sun and planets and r_9 and n_9 refer to the mean distance and motion of Neptune. This amounts to about 2·2, about 600 times the actual angular momentum of the system.

Now a condensation within the mass must have revolved in the same period as the mass itself, and thus the angular velocity of the mass when Neptune was formed must have been n_9. Hence, if k_9 be the radius of gyration at that time, we must have

$$Mk_9{}^2n_9 = 3{\cdot}7 \times 10^{-3} \qquad \text{.........(2)},$$

$$\tfrac{2}{5}Mr_9{}^2n_9 = 2{\cdot}2 \qquad \text{.........(3)},$$

whence

$$k_9{}^2/\tfrac{2}{5}r_9{}^2 = 1/600 \qquad \text{.........(4)}.$$

Thus the radius of gyration of the nebula was only $\frac{1}{25}$ of that of a homogeneous ellipsoid of the same diameter, indicating that its mass was strongly condensed towards the centre, even in the early stages. The suggestion developed by Roche (*loc. cit.*) that the sun was already formed when the condensation started is therefore essential to the theory*.

1·32. The effect of this central condensation is that the gravitation of the mass reduces practically to that of a heavy particle at the centre, and in these circumstances it is pointed out by Jeans† that there are no possible asymmetrical steady states at all; the symmetrical state is always stable, and the course of development will therefore proceed through all the symmetrical forms found by Roche until ejection commences around the equator. Hence internal condensations are impossible and the theory fails.

1·33. The argument may be presented in another way by means of a theorem due to Poincaré. If the angular velocity of the fluid in the neighbourhood of the suggested condensation is ω, consider the function

$$U_1 = U + \tfrac{1}{2}\omega^2(x^2 + y^2) \qquad \text{.........(5)},$$

where U is the gravitational potential and x and y are rectangular coordinates measured from some point within the condensation and parallel to the equatorial plane. Consider any closed surface within the condensation. Then we have by Green's theorem

$$\iint \frac{\partial U_1}{\partial n}\, dS = \iiint \nabla^2 U_1\, d\tau \qquad \text{.........(6)},$$

* The importance of the angular momentum criterion was first pointed out by Babinet (*Comptes Rendus*, **52**, 1861, 481–84) and by David Trowbridge (*Amer. Journ. of Science and Arts*, Ser. 2, **38**, 1864, 344–60). Both of these writers, however, overlooked the fact that the orbital angular momentum of Jupiter much exceeds the rotational angular momentum of the sun, and attended only to the latter. The first correct discussion is by Fouché (*Comptes Rendus*, **99**, 1884, 903–906). For an account of the primitive density distribution necessary to account for all the planets, reference may be made to Anne Sewell Young (*Astrophysical Journal*, **13**, 1901, 338–43), who undertook the work at the suggestion of Prof. F. R. Moulton.

† *Problems of Cosmogony and Stellar Dynamics*, p. 148.

the integral on the left being taken over the surface and that on the right throughout the interior. Now

$$\begin{aligned}\nabla^2 U_1 &= \nabla^2 U + 2\omega^2 \\ &= -4\pi f\rho + 2\omega^2 \qquad \text{.........(7)},\end{aligned}$$

where ρ is the density and f the constant of gravitation. If ρ_1 is the mean density inside this surface and ϕ its volume, the integral on the right is therefore equal to $2\phi\,(\omega^2 - 2\pi f\rho_1)$.

Now $-\dfrac{\partial U}{\partial n}$ is the normal component of gravity inwards across the surface, and $\dfrac{\partial}{\partial n}\,\tfrac{1}{2}\omega^2(x^2+y^2)$ is the inward normal acceleration required to keep a particle in contact with a surface rotating with angular velocity ω. Hence if

$$\frac{\partial U_1}{\partial n} = \frac{\partial U}{\partial n} + \frac{\partial}{\partial n}\,\tfrac{1}{2}\omega^2(x^2+y^2) \qquad \text{.........(8)}$$

is positive, the inward attraction due to gravity is less than is required to hold a particle in contact with the surface, and thus the combined effect of gravity and rotation will tend to make the fluid move outwards across the closed surface. In addition the fluid pressure must increase inwards, since the pressure must increase with the density, and therefore the effect of pressure is always to produce motion outwards. But for a condensation to develop it is necessary that the fluid shall have at all points an inward or a zero acceleration, and therefore $\dfrac{\partial U_1}{\partial n}$ must be negative and sufficiently great to counteract the outward acceleration due to pressure. This must be true at all points of the surface if the effect is to be a permanent condensation around a point. Hence $\displaystyle\iint \frac{\partial U_1}{\partial n}\,dS$ over the surface must be negative, and therefore $2\phi\,(\omega^2 - 2\pi f\rho_1)$ must be negative. Thus the condition for condensation to be possible is that $2\pi f\rho_1$ shall be greater than ω^2.

If the whole mass rotated like a rigid body, ω would be the same at all points, and therefore equal to n, the angular velocity of the whole nebula, and to the present mean motion of the planet formed at that stage. Now by Kepler's third law we have

$$\omega^2 = \frac{fM}{r^3} \qquad \text{.........(9)}.$$

Hence the condition for condensation is that $2\pi f\rho_1$ shall be greater than fM/r^3, or that ρ_1 shall be greater than $M/2\pi r^3$. But the angular momentum condition (4) shows that the density near the outer boundary is only a small fraction of the mean, and in this case the form of the outer boundary is such that

$$M/2\pi r^3 = 0{\cdot}36\bar{\rho} \qquad \text{.........(10)},$$

where $\bar{\rho}$ is the mean density*. Hence the condition for condensation is

* Jeans, *loc. cit.* p. 150.

that the density in the neighbourhood of the condensation when condensation begins shall be greater than 0·36 of the mean density; which contradicts the result that nearly all the mass must be near the centre, the density as far out as the orbit of Neptune being at most only of the order of a thousandth of the mean.

1·34. If we suppose condensation to have occurred in the matter that has been shed around the equator, a similar contradiction arises. Each portion of the matter in this part must revolve independently about the centre, with an angular velocity nearly equal to n, that appropriate to a circular orbit at the same distance, namely $(fM/r^3)^{\frac{1}{2}}$. Now the rotation of the matter is given by

$$2\omega = \frac{1}{r}\frac{d}{dr}(r^2 n) = \tfrac{1}{2}n \qquad \text{.........(1).}$$

Thus the condition for condensation is that ρ_1 shall be greater than $\omega^2/2\pi f$ or $n^2/32\pi f$, and finally greater than $0{\cdot}022\bar{\rho}$; much less than in the former case, but still quite inconsistent with the data of the problem. Hence internal condensation is impossible in either portion of the nebula.

1·4. *Summary*. The theory of rotational instability is therefore not a possible explanation of the origin of the solar system, and search must be made elsewhere for the correct theory. It has actually been proved by Jeans in his elaborate investigation that a nearly homogeneous mass broken up by rotational instability would give rise to a double or multiple star, the masses of the components being comparable; while a mass with a strong central condensation, if it condensed elsewhere at all, would probably give a spiral nebula, the arms consisting of streams of stars, each with a mass comparable with that of the sun. In neither case would anything resembling the solar system be produced. A gaseous body with a mass comparable with that of the sun, and strongly condensed towards the centre, could not condense at all except by absorption into the sun.

CHAPTER II

The Tidal Theory of the Origin of the Solar System

"In six days the Lord made heaven and earth, the sea, and all that in them is." Exodus xx. 11.

2·1. Long before it had been definitely proved that Laplace's Nebular Hypothesis could not give a dynamically satisfactory account of the origin of the solar system, and could not be modified to give one so long as the notion of gradual development from a nebula initially fairly symmetrical was retained, many astronomers had felt that it involved too many unproved and unplausible assumptions, and were seeking for an alternative hypothesis that would avoid the difficulties. The chief of these alternatives was the Planetesimal Hypothesis of Professors T. C. Chamberlin and F. R. Moulton, both of Chicago; this hypothesis has obtained the support of a considerable number of geologists, although astronomers have given it much less attention. Nevertheless its astronomical side appears to contain a large amount of truth, in spite of several serious objections that can be urged against it. One of these objections was pointed out by me in 1916*, but the advantages of this hypothesis over those previously advocated were so striking that I attempted to modify it in such a way as to avoid the objection. An outline of the result was published in 1917†, and various applications of the modified hypothesis to account for phenomena of the existing solar system were developed in 1918‡.

2·2. *The Methods of Rupture.* Dr J. H. Jeans has discussed on general dynamical principles the whole problem of the possible methods of rupture of fluid masses, and was able to show, in a thesis§ awarded the Adams Prize of the University of Cambridge in 1917, that the only method that could lead to anything resembling the solar system required almost the same initial conditions as the Planetesimal Hypothesis and that which I adopted. Since the two discussions rested on very different data, the close agreement between the inferences is a strong argument for the truth of some closely similar theory. There were, however, a few points of difference; these will be indicated in the following account, which differs somewhat from both presentations. The discussion of the Planetesimal Hypothesis itself will be reserved until Appendix I.

* *M.N.R.A.S.* **77**, 1916, 84–112.
† *Science Progress*, No. 45, 1917, 52–62.
‡ *M.N.R.A.S.* **78**, 1918, 424–41.
§ *Problems of Cosmogony and Stellar Dynamics*, 1919.

The fundamental feature of the hypothesis is the approach to the sun of a star considerably more massive than itself. This raised two large tides on the sun, the greatest protuberances being at the points of the sun nearest to and furthest from the star. When the distance between the two bodies became sufficiently small, the tendency to disruption due to the difference between the attractions of the star on the two opposite sides of the sun became greater than the sun's gravitation could counteract, and a portion of the sun was torn away. This afterwards condensed to form the planets and satellites.

Jeans has shown that in the tidal theory, as in the rotational theory, a mass may be broken up in two ways, according as it is approximately homogeneous or strongly condensed towards the centre. The former type would give a number of bodies with masses of the same order of magnitude; the latter is the type required to account for the solar system, and will therefore be considered in some detail.

The gravitation potential due to the sun and a star together, when the sun is supposed to have nearly all its mass concentrated near the centre and the star is supposed spherical, is

$$U = \frac{fM}{r} + \frac{fM'}{r'} \qquad \text{.........(1)},$$

where f is the gravitational constant, M and M' are the masses of the sun and star, and r and r' the distances of the point considered from the centres of the respective bodies. If coordinates x, y and z be taken, the origin being at the centre of the sun and the x axis pointing towards the star, and if u, v, w be the component velocities of the matter of the solar envelope at any point, we have

$$\frac{d}{dt}(u, v, w) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) U - \frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) p \text{ ....(2)},$$

where the derivatives with regard to the time indicate the accelerations in the directions of the axes, p is the pressure and ρ the density. Now the acceleration of the central body towards the star will be fM'/R^2, where R is the distance between the centres. Hence if (u_0, v_0, w_0) be the velocity of the centre of the sun,

$$\frac{d}{dt}(u_0, v_0, w_0) = \left(\frac{fM'}{R^2}, 0, 0\right) \qquad \text{.........(3)},$$

and if (u', v', w') be the velocity of a particle of the envelope relative to the centre of the sun, we have, since R is independent of x, y and z,

$$\begin{aligned}\frac{d}{dt}(u', v', w') &= \frac{d}{dt}(u - u_0, v - v_0, w - w_0)\\ &= \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\left(U - \frac{fM'x}{R^2}\right)\\ &\quad - \frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) p \qquad \text{.........(4)}.\end{aligned}$$

The discussion of the resulting motion may be facilitated by considering two extreme cases, called by Jeans 'slow' and 'transitory' encounters. In a slow encounter, the changes are slow compared with the corresponding free vibration of the solar envelope, while in a transitory one they are quick.

The changes during a slow encounter bear a certain resemblance to the motion of a pendulum when the bob is drawn aside by a gradually increasing horizontal force. It moves slowly in the direction of the force, the deflexion at any instant being almost the same as if the pendulum were in equilibrium under the action of gravity and the force actually acting at that instant. In a slow tidal encounter, approximate equilibrium is similarly maintained, and the form of the solar envelope can be calculated as if it were in equilibrium under the action of the gravitation of the sun and the star. Thus the acceleration $\frac{d}{dt}(u', v', w')$ can be neglected.

In the case of the transitory encounter, the analogous pendulum problem is when the bob is suddenly struck by a horizontal blow, and then allowed to swing or revolve freely. In the tidal problem, the motion is as if a motion were suddenly imparted to the envelope, and this were then left to readjust itself under the mutual attraction of its parts and the sun's gravitational field.

2·21. In a slow encounter we have

$$\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\left(U - \frac{fM'x}{R^2}\right) = \frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) p \quad \text{.........(1).}$$

Thus the fluid pressure and the density must be functions of $U - \frac{fM'x}{R^2}$, that is, of $\frac{M}{r} + \frac{M'}{r'} - \frac{M'x}{R^2}$. Jeans (*loc. cit.* pp. 153–56) denotes this latter quantity by Ω. Thus the outer boundary of the mass will be one of the surfaces where Ω is constant; it will be the particular surface of the set

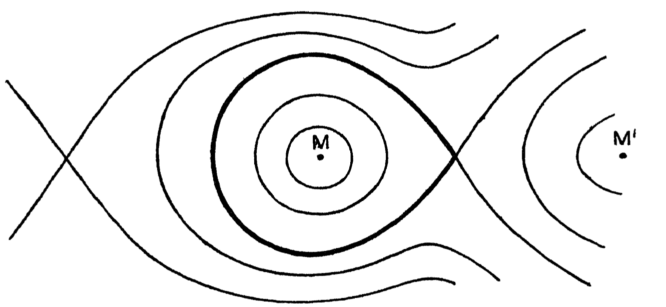


Fig. 1. (From Jeans's *Problems of Cosmogony.*)

that is just large enough in volume to be filled by the matter of the solar envelope. Jeans shows that one of these surfaces, represented by the thick curve in Fig. 1, surrounds the sun completely and has a sharp point

on the side nearest the star. At first the volume of the solar envelope is much less than that of this surface; but as the star approaches, the linear dimensions of this critical surface diminish in proportion to the distance between the two bodies, until the volume becomes too small to hold the envelope. The critical value of R, if the disturbing body has a mass double that of the sun, is $2{\cdot}87a$, where a is the undisturbed radius of the envelope. When this state is reached $\frac{\partial \Omega}{\partial x}$ vanishes at the apex; and this implies that whereas up to this stage everywhere, and even at this stage everywhere else, the resultant influence of gravity is to accelerate the matter towards the interior, this influence now ceases at this point. With a closer approach it changes its direction, and the matter of the envelope is drawn out at the point in the direction of the star. The action of fluid pressure is always to encourage ejection, which will therefore continue until the star has receded again to such a distance that its gravity is no longer enough to neutralize that of the sun.

Initially the velocity and acceleration of the ejected matter are both inappreciable, but when the star approaches closer the acceleration increases and the matter therefore begins to move straight towards the star. Meanwhile the star is moving transversely as well as towards the sun; thus, when the matter has moved some distance outwards, the star is no longer in its direction of motion, and will therefore attract it sideways. Thus a velocity around the sun as well as away from it will be acquired. The changes in the condition of the system are indicated in the following diagram, where the dotted curve shows the position of the solar envelope at the first rupture and the figures 1, 2, and 3 indicate the positions and paths of the particles ejected at the first, intermediate, and final stages. Since all the accelerations relative to the sun are in one plane, namely that of the motion of the star relative to the sun, it follows that the whole of the motion produced must be approximately in this plane, apart from a possible slight departure due to the rotation of the sun before the passage of the star. Further, the transverse motion of every portion of the ejected matter is in the same sense as the motion of the star. It would be incorrect, however, to suppose that the whole of the ejected matter would follow, even at first, the same orbit. The point where emission is taking place is necessarily immediately towards the star, and therefore is continually changing as the star moves, its path relative to the centre of the sun being geometrically similar to the relative path of the star. Thus all particles start from different positions and therefore travel on different paths; they do not follow one another along one path, as might at first be thought.

If ejection was continuous, all the ejected matter would at any subsequent time lie within a long narrow region; in other words, it would form a filament. In shape this would resemble a boomerang. For, subject

to the conditions of a slow encounter the velocity of the matter relative to the sun is inappreciable before ejection, and the acceleration is zero at the moment of ejection. Hence the end of the filament nearest the centre must touch the locus of the equilibrium point, while the acceleration away from the sun as the star approaches will produce a considerable outward velocity, so that the filament will be more strongly curved away from the sun than the locus of the point of ejection.

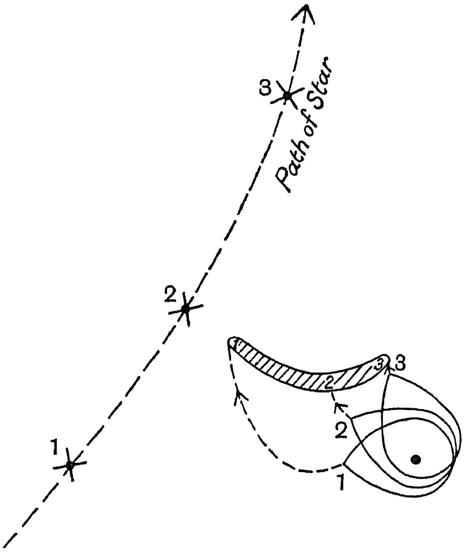


Fig. 2. Diagram of the changes in the form of the solar envelope and the paths of portions of the ejected matter during the passage of the star.

It will be seen that when the star reaches its nearest point to the sun, the point of ejection is also at its nearest. When the star recedes in the least, the sun's attraction on the matter severed at the moment of closest approach will again be greater than the tidal disturbing force, and it will therefore draw this matter back into the solar envelope. We see similarly that no ejection of matter not already separated can take place during the retreat of the star, and indeed that what has been ejected during the approach will fall back, with the exception of such as has acquired a sufficient transverse velocity to miss the envelope when it approaches the sun on its next revolution. Thus only the parts first drawn off can remain permanently detached from the sun. At any stage of ejection the volume of the critical surface would be decreasing at a rate simply proportional to the rate of change of volume of a sphere with its centre at the sun and extending to the star. Thus when ejection started the rate of ejection would be finite if the rate of approach was finite, would increase to a maximum, and would fall to zero at periastron*.

* This account differs from that of Jeans, *loc. cit.* p. 283: "The rate of ejection of matter would be slow at first, would increase to a maximum when the passing star was at its nearest to the sun and would subsequently diminish to zero."

The typical slow encounter just described is, however, incapable of being realized in actuality, for quantities supposed negligible in it are in fact incapable of being small. The period of a free vibration of tidal type in the solar envelope could hardly differ in order of magnitude from the period of revolution of a planet moving in a circular orbit at the point of ejection. This period is $2\pi\,(b^3/fM)^{\frac{1}{2}}$, where b is the radius to the point of ejection. The time of passage of the star, even if the relative velocity before it was appreciably affected by the sun was inappreciable, would be of the order of $\pi\,\{R^3/f\,(M+M')\}^{\frac{1}{2}}$, by the ordinary theory of two bodies. Now the condition for any emission to occur is approximately that b shall exceed $(M/2M')^{\frac{1}{3}}R$. Thus the ratio of the two times considered is of order $\left(\dfrac{2\,(M+M')}{M'}\right)^{\frac{1}{2}}$, which cannot be less than unity, whereas it would have to be a small fraction for the conditions of a slow encounter to be satisfied. Any actual encounter would be intermediate in character between the typical slow encounter and the typical transitory one; but whereas the former type cannot be even approximately realized, being dynamically impossible for freely moving bodies, the latter can be realized with any required degree of accuracy.

2·22. Consider now a transitory encounter. The equation of continuity is

$$-\frac{1}{\rho}\frac{d\rho}{dt}=\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z} \qquad \text{.........(1)},$$

where the axes are now taken to be fixed in direction and not accelerated, and u, v, w are the component velocities of the fluid. Now by hypothesis the encounter is such as will make the velocities moderate*, while its duration is inappreciable. Hence the quantity on the right is moderate, and therefore its integral for the whole or any part of the encounter is inappreciable; thus ρ, the density of any element of fluid, does not change during the encounter.

If the fluid be compressible, the pressure on any element of the fluid is determinate when the density is known, and therefore, by the last paragraph, does not change during the encounter. Now the accelerations of any element are given by

$$\frac{d}{dt}(u, v, w)=\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)U-\frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)p \qquad \text{.........(2)}.$$

Let us denote $\int U dt$, taken for the whole duration of the encounter, by

* In arguments based chiefly upon the notion of orders of magnitude, some set of quantities must be taken as a standard. Other quantities are called 'large,' 'small' or 'moderate' in comparison with these. The standard quantities must, of course, be consistent in the same sense as is applied to units of measurement. Thus, in the present problem the mass of the sun, gravity at the surface of the primitive sun, the velocity and mean motion of a planet at the sun's surface and moving about it in a circular orbit, and the radius of the sun, are all standard quantities, while the pressure at the sun's centre is moderate.

Φ. By the last paragraph, the second vector on the right is essentially moderate, and its integral throughout the encounter is therefore negligible. Thus if we integrate the equations of motion for the duration of the encounter we shall have

$$u, v, w = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\Phi \qquad \text{.........(3).}$$

Thus there is no impulsive pressure during a transitory encounter.

If the fluid be incompressible, the argument breaks down, but the result may be obtained otherwise. Equations (2) still hold, but, ρ being constant for any given element, we can denote $\int p\,dt$ for the duration of the encounter by ϖ, and then if we integrate (2) for the duration, we have

$$(u, v, w) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\Phi - \frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\varpi \qquad \text{.........(4).}$$

Differentiating the three equations thus combined with regard to x, y, and z respectively, and adding, we have

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = \nabla^2\Phi - \left\{\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial z}\right)\right\} \text{...(5).}$$

Now the left side is equal to $-\frac{1}{\rho}\frac{d\rho}{dt}$, which is zero since the fluid is incompressible. $\nabla^2 U$ is equal to $-4\pi f\rho$, which is always moderate, and therefore $\nabla^2\Phi$ is inappreciable. Hence

$$\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial z}\right) = 0 \qquad \text{.........(6).}$$

If we had an uncharged dielectric with a specific inductive capacity proportional to $\frac{1}{\rho}$, this differential equation would be satisfied if ϖ were the electrostatic potential within it. Further, the pressure, and therefore ϖ, vanish at all points of the boundary of the envelope. The problem is therefore completely analogous to that of an uncharged dielectric with a zero potential over the boundary, which is known to have a unique solution*, namely that the potential is zero for all points within it. Hence we have the general result that there is no impulsive pressure in a transitory encounter, and therefore equations (3) hold in all cases.

2·221. Now by hypothesis the gravitation of the sun is inappreciable in comparison with that of the star during the encounter, even within its envelope, and *a fortiori* it is therefore insufficient to divert the path of the star appreciably. The star will therefore move in a straight line. Taking the path of the star to be the axis of y and its velocity to be v, the coordinates of the star are $(0, vt, 0)$ and the accelerations of a particle (x, y, z) due to the star are

$$-fM'\left(\frac{x}{r'^3}, \frac{y - vt}{r'^3}, \frac{z}{r'^3}\right),$$

where r' is the distance of the particle from the star.

* Cf. Jeans, *Electricity and Magnetism*, 1908, 161–62.

Integration of these with regard to the time gives the velocity of the particle at the end of the encounter. Putting $x^2 + z^2$ equal to $R_1{}^2$, so that R_1 is the least distance of the star from the particle considered, and $y - vt$ equal to $R_1 \tan \phi$, we have

$$r'^2 = x^2 + z^2 + (y - vt)^2$$
$$= R_1{}^2 \sec^2 \phi,$$

and
$$- v\,dt = R_1 \sec^2 \phi\, d\phi.$$

The direction of the final velocity evidently intersects the path of the star. The y component is
$$\frac{fM'}{R_1 v} \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \sin \phi\, d\phi,$$

which vanishes. The component at right angles to the path of the star is

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{fM'}{R_1{}^3 \sec^3 \phi} \cdot R_1 \cdot \frac{R_1}{v} \sec^2 \phi\, d\phi = \frac{2fM'}{R_1 v}.$$

If R_0 be the distance of closest approach to the centre of the sun, and if

$$R_1 = R_0 - a,$$

the velocity of the particle relative to the centre of the sun at the end of the encounter will be

$$\frac{2fM'}{v} \left(\frac{1}{R_0 - a} - \frac{1}{R_0} \right) = \frac{2fM'a}{R_0{}^2 v} \qquad \text{.........(1)}$$

approximately.

The condition that the sun may be broken up is evidently that this velocity shall be comparable with the parabolic velocity for a particle at the outside of the envelope. If a be the mean radius, this condition becomes

$$\frac{2fM'a}{R_0{}^2 v} > \left(\frac{2fM}{a} \right)^{\frac{1}{2}} \qquad \text{.........(2).}$$

We have the further condition that the duration of the encounter shall be less than the period of oscillation of the mass. The former can be taken to be $2R_0/v$, and the latter half the period of revolution of a planet at the boundary. Hence

$$\pi \left(\frac{fM}{a^3} \right)^{-\frac{1}{2}} > \frac{2R_0}{v} \qquad \text{.........(3).}$$

Multiplying (2) and (3), we find on simplification

$$\sqrt{2}\, \frac{M'}{M} > \frac{R_0{}^3}{a^3} \qquad \text{.........(4).}$$

The corresponding condition found for a slow encounter was equivalent to

$$\frac{2M'}{M} > \frac{R_0{}^3}{a^3} \qquad \text{.........(5).}$$

Thus for a given periastron distance a larger disturbing body is required to produce rupture in a transitory encounter than in a slow one. The difference is not, however, great.

For a given mass M' and a given minimum separation of the two bodies, equation (2) imposes an upper limit to the values of v consistent with rupture taking place. To indicate the possible dimensions of the system at rupture, take

$$M' = 4M, \; R = 1{\cdot}5a, \; v = 4 \times 10^6 \text{ cm./sec.}$$

With $$M = 2 \times 10^{33} \text{ gm.}, \; f = 6{\cdot}6 \times 10^{-8} \text{ c.g.s.},$$

(2) gives $$a < 6{\cdot}4 \times 10^{13} \text{ cm.},$$

which is the present mean distance of a rather remote asteroid.

The extreme type of transitory encounter may be approached indefinitely closely by supposing M' and v increased indefinitely while retaining a constant ratio to each other. It is easy to see, however, that this type of encounter could not lead to anything resembling the solar system. The investigation just given shows that in a purely transitory encounter the motion produced necessarily consists of the shooting out of two protuberances, one towards the star and one away from it, but both directly away from the centre of the sun. Some of the matter might acquire enough velocity to make it leave the sun's influence altogether; but such as did not would fall straight back into the sun and be reabsorbed. Thus no planetary system would be generated. To account for the actual solar system an encounter of intermediate type is required.

2·3. *The Density of the Primitive Sun.* So far the only postulate we have made about the density of the primitive sun is that it was not nearly uniform; it must have been strongly condensed towards the centre. The sun at present has a density rather greater than that of water, and it seems unlikely on several grounds that a dwarf star could have been as heterogeneous as the present theory requires. Thus the origin of the planets must have happened while the sun was in the giant stage. It was then a gaseous star, and the structure of gaseous stars has been the subject of much investigation. The temperature must have been sufficiently high to keep the constituents of the present planets in a gaseous state; and if we take silicon and the less fusible metals as our standard, as we must if we are to explain how the earth came to be formed from the sun, the effective temperature of the primitive sun cannot have been under 3000° absolute. E. A. Milne, using Eddington's* theory of the constitution of gaseous stars, has shown that a gaseous star of the mass of the sun, and with an effective temperature of 3100°, would have a mean density of 5×10^{-5} gm./cm.^{3}, and a radius of $2{\cdot}1 \times 10^{12}$ cm., or 0·14 astronomical unit. This radius is much less than that of Betelgeuse, which is comparable with 1·5 astronomical units, the difference being attributable to the much greater mass of the latter star. If the temperature was higher, the density would be greater and the radius correspondingly less. Thus the above

* *M.N.R.A.S.* 77, 1916–17, 16–35, 596–612.

estimate is to be regarded as a lower limit to the possible density of the primitive sun.

2·4. *The Rupture of the Filament.* Whether the encounter was slow or transitory, the result must therefore have been the formation of a long protuberance towards the star; transverse motion, if any, would be in the plane of the star's relative motion. In either case it is possible, though not necessary, for a shorter protuberance to be formed on the side away from the star; this depends on the mass of the star and on its distance of closest approach. In the slow encounter the filament would be formed by ejection of matter from a limited region of the boundary; the shorter filament, if formed, would be diametrically opposite to it. In a transitory encounter the whole envelope would commence to be stretched, the outward velocity relative to the central body being greatest on the side nearest to the periastron, a smaller outward velocity being also produced on the opposite side. The actual encounter would be of intermediate type. The longer protuberance produced will in general be called 'the filament.'

We have now to consider the subsequent development of the system. As the star retreated, much of the matter drawn out, including perhaps the whole of the shorter protuberance, must have fallen back into the sun. The outer portions of the longer one would, however, continue to recede from the sun, and, being deflected transversely by the star, would miss the sun on the return journey and continue to revolve around it. There is no necessary upper limit to the length of a filament produced in this way; if the initial velocity was greater than the parabolic velocity at the point of emission, gravity would be unable to prevent the first matter drawn out from receding to an infinite distance, and the fluid pressure of the matter behind it would always be pushing it onwards. Thus we should expect a long narrow filament to be formed.

If such a strip were steadily drawn out under the joint influence of the sun and the star, the mass per unit length in any region of it would presumably vary steadily with the distance from the sun, having possibly one maximum in the middle. If, however, it received a slight distortion, the mass per unit length in some region being increased, it is possible that the extra gravitative power of this region would draw other matter towards it. The disturbance would then increase exponentially with the time, and therefore a condensation would be formed in the filament.

The theory of the aggregation of a gaseous filament in this way has been outlined by Jeans (*loc. cit.* pp. 157–60). Condensation would begin when the length of the strip already ejected reached the value

$$l = \tfrac{1}{2} c \left(\frac{\pi}{f\rho}\right)^{\frac{1}{2}} \qquad \text{.........(1)},$$

where c is the velocity of sound in the gas. The ejected portion would then begin to detach itself from the main body; thenceforth it would lead an

independent existence as a planet. If ejection continued, a second planet would be formed, and others would follow until ejection ceased. Those formed last would fall back into the sun as the disturbing body receded, but the early ones would retain their identity.

The above condition is necessary for condensation, but not sufficient. Two factors will tend to make a gaseous filament dissipate itself instead of condensing; namely, first, the tendency of a gas to spread itself out into the surrounding vacuum, and second, the tidal disruptive action of the sun itself. A gas suddenly liberated into a vacuum will spread out with a velocity equal to a small multiple of the velocity of sound in the gas. According to the usual, but probably inaccurate, formula*, the velocity of efflux of a gas into a vacuum should be $\left(\frac{2}{\gamma-1}\right)^{\frac{1}{2}}c$, where γ is the ratio of the specific heats. For a monatomic gas this amounts to $\sqrt{3}c$. The condition that the filament should not dissipate itself is that this must be less than the velocity necessary to remove a particle from the gravitative influence of the filament. It is, however, more convenient mathematically to consider the corresponding criterion for a primitive planet. If a be the radius of such a planet, and m its mass, this condition shows that

$$3c^2 < 2fm/a \qquad \text{.........(2).}$$

If b be the radius of the filament and we assume that the density did not change considerably while a section was collecting into a nearly spherical form, we have

$$m = \pi\rho b^2 l = \tfrac{4}{3}\pi\rho a^3 \qquad \text{.........(3),}$$

and therefore

$$3c^2 < \tfrac{8}{3}\pi f\rho a^2 \qquad \text{.........(4).}$$

But from (1) we have

$$c^2 = \frac{4}{\pi} f\rho l^2 \qquad \text{.........(5),}$$

and from (4) and (5) we find

$$a > \frac{3}{\pi\sqrt{2}}\, l \qquad \text{.........(6).}$$

But

$$\tfrac{3}{4} b^2 l = a^3 > \frac{27}{\pi^3 2^{\frac{3}{2}}}\, l^3 \qquad \text{.........(7),}$$

whence

$$b > 0{\cdot}5l \qquad \text{.........(8).}$$

Hence the length and thickness of a section of the filament at rupture would be comparable, and there is no reason to suppose that sufficiently great changes took place in the ratio of its longitudinal and transverse dimensions to invalidate the supposition that the order of magnitude of the density remained unaltered during the adoption of the nearly spherical form.

Now substituting for a and l in (3) we find in turn

$$m > \frac{2^{\frac{1}{2}}3^2}{\pi^2}\,\rho l^3 > \frac{3^2}{2^{\frac{5}{2}}\pi^{\frac{1}{2}}}\,\frac{c^3}{f^{\frac{3}{2}}\rho^{\frac{1}{2}}} \qquad \text{.........(9).}$$

* Ramsey, *Hydromechanics*, Part II, 1913, 59.

This gives a lower limit to the mass of a planet that could be formed by gradual condensation of a gaseous filament. For nitrogen at 273° absolute the value of c is about 3×10^4 cm./sec. It is proportional to the square root of the absolute temperature and to the inverse square root of the molecular weight. Monatomic silicon would have the same molecular weight as diatomic nitrogen; thus c for silicon at 3000° absolute would be $3 \times 10^4 \left(\frac{3000}{273}\right)^{\frac{1}{2}}$ or practically 10^5 cm./sec. The mean density of the sun has been seen to have been at least 5×10^{-5} gm./cm.^{3}, and would be greater if the sun were more condensed. On the other hand, the sun would be condensed towards the centre, and the ejected matter would be less dense than the average. As a rule it will be supposed as a working hypothesis that the density of the ejected matter was 5×10^{-5} gm./cm.^{3}. Then

$$m > 0{\cdot}8 \times 10^{28} \text{ gm.} \qquad \text{.........(10)},$$

$$a > \left(\frac{9c^2}{8\pi f\rho}\right)^{\frac{1}{2}} > 3 \times 10^{10} \text{ cm.} \qquad \text{.........(11)}.$$

The radius of the primitive planet comes out at about $\frac{1}{60}$ of that of the primitive sun, indicating that planets with a larger radius than this could condense. It is likely that the thickness of the filament, even in the nearest realizable conditions to a slow encounter, would be a moderate fraction of the radius of the sun, so that this result indicates that the formation of planets by condensation of the filament is possible. The mass indicated is, however, intermediate between those of the great planets and of the terrestrial planets. But in a giant star the law that the luminosity is a function of the mass alone shows that the absolute effective temperature must have been proportional to the inverse square root of the radius. Therefore c is proportional to the inverse fourth root of the radius, and $c^3\rho^{-\frac{1}{2}}$ is proportional to the three-fourths power of the radius. Thus if the sun was more condensed than has so far been supposed, the primitive planets could have been less massive than has just been shown possible.

2·41. There is, however, a definite lower limit to the possible masses of planets produced by gradual condensation from matter ejected from a gaseous star. For, returning to equation 2·4 (2),

$$3c^2 < 2fm/a \qquad \text{.........(1)},$$

and supposing that when the planet ultimately solidified its radius was a' and its density ρ', we have

$$a > a' \qquad \text{.........(2)},$$

since the body must contract in passing from the gaseous to the solid state. Hence

$$3c^2 < 2fm/a' \qquad \text{.........(3)}.$$

If the density of the solid planet be taken at the reasonable value of 3 gm./cm.^{3}, m is nearly $12a'^3$, and therefore

$$24fa'^2 > 3c^2 \qquad \text{.........(4)},$$

which with the data so far adopted gives

$$a' > 1{\cdot}4 \times 10^8 \text{ cm.} \qquad \text{.........(5).}$$

Thus no planet with a smaller radius than 1400 km. could have been formed by gradual condensation from the gaseous state. If the primitive sun was hotter, c would be greater, and this limit would have to be raised.

This result is not dependent on any particular theory of the origin of the solar system, and the existence of many bodies within the system whose sizes fall below this limit establishes, as decisively as anything can be established in cosmogony, that these bodies were not formed by slow condensation from the gaseous state. They include all the asteroids, and all the satellites except the moon, Titan and the four great ones of Jupiter. Thus at least two processes must have been operative in the formation of the planets and their satellites.

2·5. *The Condensation of the Planets.* The cooling of a condensation as massive as the great planets can now be described in some detail. Eddington has shown* that in a giant star a column of matter containing one gram per square centimetre of cross section would absorb all but the fraction $e^{-5\,4}$ of the radiation falling normally on its end. With the above data, radiation from the centre to the circumference would have to traverse $1{\cdot}5 \times 10^6$ grams per square centimetre before reaching the outside, and therefore, unless the opacity dropped to a ten-millionth or less of its previous value as soon as it was ejected, it is justifiable to regard the radiation emitted by the filament and the primitive condensations as coming from a surface layer.

Imagine a sphere of radius a and effective temperature V; let its mean density be ρ and let us denote Stefan's constant by σ. Then the total rate of loss of energy by radiation is $4\pi\sigma a^2 V^4$, and the rate of loss per gram of matter is $3\sigma V^4/\rho a$. Taking

$$\begin{aligned} \sigma\,(1^\circ)^4 &= 5 \times 10^{-5} \text{ ergs/cm.}^2 \text{ sec.,} \\ V &= 3000^\circ, \\ \rho &= 5 \times 10^{-5} \text{ gm./cm.}^3, \\ a &= 3 \times 10^{10} \text{ cm.,} \end{aligned}$$

the rate of loss of energy is 8000 ergs per gram per second, or $\frac{1}{5000}$ calorie per gram per second. Thus the matter would be losing heat at the rate of 6000 calories per gram per year, which would be enough to ensure its liquefaction in a few revolutions about the sun, and possibly before it had completed its first revolution, were cooling not delayed by internal atomic changes, energy set free by contraction, and solar radiation. All of these would be less effective in a gaseous planet than in the sun, and as they were in approximate equilibrium with radiation before the disruption, radiation would overwhelm them after it.

* *M.N.R.A.S.* **77**, 1917, 602.

The mutual gravitation of the parts of a great planet would therefore hold it together, while radiation from the surface would gradually liquefy it. Since cooling would take place at the outside, drops would be formed there and would fall inwards under gravity. They would collect with the densest at the centre; solidification would in due course set in and proceed till complete.

2·51. Smaller masses of the same original density, or less dense masses of the same linear dimensions, would have a much more complicated history. In this case the attraction of the filament or of the condensation would be inadequate to retain the outer portions, which would therefore spread out with a velocity not exceeding $\sqrt{3}$ times the velocity of sound. Radiation would continue simultaneously. If A denote the surface of a segment, the rate of loss of energy by radiation is $\sigma A V^4$. The temperature would fall to the boiling point by radiation still more quickly than in the case of a larger mass. Further radiation would procure liquefaction, and a volume of gas equal to $\sigma A V^4/\rho L$ would therefore liquefy in unit time, where L is the latent heat of evaporation. Meanwhile the rate of increase of volume through spreading would be $\sqrt{3}Ac$. If the former of these was the greater, the reduction in volume through liquefaction would leave sufficient space to accommodate the expanding gases. The mass would therefore condense into liquid drops near the outside, and the still uncondensed gases would spread so as to fill only partially the space left by condensation. Thus expansion in volume would cease. The condition for this is that ρ shall be less than $\sigma V^4/\sqrt{3}Lc$, which with our previous data, taking L equal to 600 calories per gram, or 24×10^9 ergs per gram, is $1{\cdot}3 \times 10^{-5}$ gm./cm.^{3}. If the matter was originally more dense than this, and gravity was insufficient to control its expansion, it would expand approximately adiabatically until it reached the critical density, when it would proceed to cool principally by radiation. If it cooled to the boiling point during the adiabatic expansion, much of it would have liquefied before it reached the critical density; if it had not cooled so far, it would liquefy at this density.

The drops formed would be at the same temperature as the gas that produced them. Now the velocity of agitation in a monatomic gas is $(\frac{9}{5})^{\frac{1}{2}}c$. The velocity of efflux is therefore 1·3 times the velocity of agitation. Thus the matter liquefying must be composed of molecules moving both inwards and outwards, for if they were all moving outwards their velocities could not differ enough among themselves to give the actual velocity of agitation. The effective front of the matter spreading out is composed chiefly of molecules moving forwards, and therefore the condensation must take place some distance behind the effective front; the latter goes straight on and is lost to the planet. Now the mass velocity of a drop after liquefaction would be the mean of the previous outward velocities of all its molecules,

and since many of these would be negative, the drops would move outward with velocities much less than that of the effective front. If their velocity was still too great for them to be retained by the gravitation of the planet, they would pass off and be lost; but their inertia would delay the expansion of the gas following them, and thereby reduce the rate of expansion of the whole. It appears probable that the rate of expansion would ultimately be reduced to a very small fraction of the velocity of sound, and the condensation would become a system of liquid particles in an approximately stationary gaseous medium. The details of the process would be very difficult to follow out theoretically, but it appears possible that the velocity of expansion might be reduced to such an extent that gravity could control it. Thus radiative cooling leads to much the same result as adiabatic cooling would; the relation between them is that adiabatic cooling is the more important when the density is greater than the critical density, and that if the density ever falls to the critical value, or if it was initially below it, radiative cooling then takes the more prominent part.

2·52. In any mode of aggregation a planet would pass through a liquid stage. For the drops would be at the boiling point to start with, and would fall in towards the centre through a mass of gas at a temperature still far above their own; for the temperature of the gas would necessarily increase inwards on any theory of its thermodynamics. Thus their temperature would be raised both by conduction and by friction. When they reached the centre they would come to rest, being further heated by their mutual impacts, and would form a liquid core. Thenceforth the core would continue to be heated by contact with the surrounding gas. This would partially condense on the core, which would itself be thereby heated to a temperature not greater than the boiling point at the actual pressure. The outer gases might either condense, the drops falling in to augment the nucleus, or else pass off into space; but it is certain that any matter aggregated into the nucleus would be liquid.

The last point is of capital importance in the development of geophysical theory, and therefore merits further consideration, which may conveniently be given here. It has been asserted by Chamberlin and others* that the drops would cool not only to the liquid but to the solid state, and that the nucleus of a small planet would accordingly have been initially and permanently solid. The objection is that a liquid particle falling through gas would necessarily be heated; if therefore it was to cool further it would have to emerge from the region occupied by gas. Thus it would have to continue to move outwards with a velocity greater than the velocity of expansion of the gas, which by hypothesis is itself too great to be controlled by gravity. Thus such a particle would be lost to the condensation. If on the other hand the planet was of such mass that it could

* See Appendix I.

control particles moving at the surface with the velocity of efflux of a gas into a vacuum, it is certain that liquid particles would commence to fall inwards as soon as they were formed. Thus in the case of a large planet the formation of a liquid core is certain; in the case of a lighter planet the alternatives are either a planet formerly fluid or no planet at all.

It has also been thought by Chamberlin and others that adiabatic cooling would proceed after condensation and continue till the mass was solid and cold. This is impossible, by what has been said above, since the surroundings would always be hotter than the core. But even if free cooling of the drops was possible, they would still not be cooled to the solid state by adiabatic cooling. For adiabatic cooling below the boiling point can be caused only by evaporation, and therefore could not lower the temperature below a point where the vapour pressure is insignificant. Thus the temperature could never be reduced in this way by more than 200° at most below the boiling point. But the difference between the boiling points and melting points of silicon and heavy metals is at least several hundred degrees. Hence adiabatic cooling could reduce the substance of the planets at most to the liquid state and not to the solid state. If evaporation was appreciable in substances with high melting points when fused, the casting of iron and the melting of glass should produce a great amount of distillation of iron and glass on to the factory walls, which does not take place.

2·53. Returning now to the question of the disruptive action of the sun we see that it could not impede condensation. For it was virtually taken into account by Jeans in finding the condition that the filament might break up into detached masses, and the length of the segments obtained by him is practically determined by the condition that the matter may be far enough from the sun not to be broken up further. As it receded from the sun, it would become more nearly spherical and more dense, and on both grounds the sun's disruptive influence would diminish, apart from the direct effect of distance.

2·54. In this way a number of liquid planets of varying sizes would be formed. They would move in orbits about the sun, all in one direction and approximately in one plane. Their outward motion would not in general be annulled during the passage of the star, so that after the encounter they would have considerable velocities away from the sun; in other words, their orbits would be highly eccentric.

2·6. *Buffon's Theory.* At this stage we may quote the passage from Laplace's *Système du monde* that was omitted on p. 6. It is as follows:

Buffon is the only person I know who, since the true system of the world was discovered, has attempted to find out the origin of the planets and satellites. He supposes that a comet, falling upon the sun, drove from it a

torrent of matter, which united far away into several globes of various sizes and at various distances from that body. These globes are the planets and satellites, which, by cooling, have become opaque and solid.

This hypothesis satisfies the first of the five conditions already mentioned*; for it is clear that all the bodies thus formed must move nearly in the plane that included the centre of the sun and the path of the torrent of matter that formed them. The four other phenomena appear to me inexplicable by this means. In truth, the absolute movement of the molecules of a planet must be in the direction of the motion of its centre of gravity, but it does not follow that the rotation of the planet will be in the same sense. Thus, the earth could turn from east to west, and yet the absolute movement of each of its molecules could be from west to east. This applies also to the motion of revolution of the satellites, whose direction, on this hypothesis, is not necessarily the same as that of the motion of their primaries.

The smallness of the eccentricities of the planetary orbits is not only very difficult to explain on this hypothesis, but actually contrary to it. We know from the theory of central forces that if a body, moving in a closed orbit about the sun, touches the surface of that luminary, it will return there in every revolution. Hence it follows that if the planets had been primitively detached from the sun, they would touch it at each revolution, and their orbits, instead of being circular, would be very eccentric. It is true that a torrent of matter expelled from the sun cannot be exactly compared to a globe grazing the surface: the pressure and the gravitational attraction between the parts of the torrent may change the directions of their movements and make their perihelia recede from the sun. But their orbits must remain permanently very eccentric; or, at least, they could have had small eccentricities only by the most extraordinary chance. Finally, one sees no reason, on Buffon's hypothesis, why the orbits of the comets already observed, about ninety in number, are all very much elongated; this hypothesis is therefore far from satisfying our conditions.

If in Buffon's hypothesis we replace 'comet' by 'star more massive than the sun' and 'falling upon the sun' by 'approaching very close to the sun,' we have a hypothesis with a close resemblance to the one that has been elaborated in this chapter. We must, therefore, be prepared to meet Laplace's objections. We note that his second point is certainly false, eight retrograde satellites being now known; and the third point is untrue of Uranus and probably of Neptune. The fifth remains a difficulty, comets not having yet been satisfactorily included in any cosmogonical hypothesis (Laplace's own being no exception). The fourth point has now been met†, and it appears that the present small eccentricities of the planetary orbits are perfectly consistent with the tidal theory.

2·7. It will be noticed that in the primitive sun the lightest materials would be in the outer layers; they would therefore be the first ejected, and would proceed to the greatest distances from the sun. Thus the outermost planets would be expected to be composed of the lightest materials and to have the lowest densities, as they have. If any matter was expelled from the solar system altogether, as is quite possible, it would be the

* See 1·1. † See Chapter IV.

lightest of all. The effect would, however, have been intensified by the fact that the lighter planets would have lost much more of their lighter materials, and would thus have come to contain a greater proportion of heavy constituents than they originally possessed.

It has been seen that much expelled matter must have fallen back into the sun, taking with it the angular momentum that it acquired during its journey. Hence the sun acquired a rotation, the plane of its equator being near the planes of the motions of the planets, as it actually is.

2·8. *The Origin of Satellites.* It appears unlikely that all the bodies in the solar system were produced in the disruption already discussed. The diameter and density of the filament would certainly vary from point to point, but it is incredible that they could vary in such an irregular way as to account for the occurrence of bodies with such widely different masses as Saturn and its satellites in close proximity and between two other bodies both comparable in mass with Saturn. This is to take only one example of the difficulties presented by the wide differences in mass between the outer planets and their satellites. A natural suggestion is that the satellites were formed from the planets. They could not have been formed by gradual condensation of the planets, for every argument used against the Laplacian theory of the origin of the planets and against its modifications is equally applicable to the corresponding theories of the origin of the satellites from the planets. The tidal theory, on the other hand, is applicable to this problem also. Each nucleus would pass near the sun at its first perihelion after it was formed. How near to the sun it would pass presents a perhaps quite intractable question in the Problem of Three Bodies; but it is at least plausible that the perihelion distance would be less than the distance of its centre from the sun when it first took a spherical form. The latter distance is specified by the condition that the body could just hold itself together in spite of the disruptive tidal action of the sun. If it had not condensed appreciably during its first revolution it would therefore be broken up by solar tidal action at its first perihelion. If it approached more closely it might be broken up even in spite of the condensation. A filament would then be produced by the planet, and would go through a process of development similar to that of the filament ejected by the sun. This is the mode of origin of the satellites suggested by Jeans. It is not, however, a complete account. Just as the planets necessarily move in one direction around the sun, any satellite would necessarily move in that direction about its primary if it had been produced in this way and always remained with its primary. But just as in the first great disruption much matter might have acquired a great enough velocity to expel it from the sun's influence altogether, so in these minor disruptions some satellites formed might have been permanently lost to their primaries and proceeded to describe independent orbits about

the sun. It is possible that in their later development they would be captured* by their own parents or by other planets, or they might remain permanently independent.

If, however, a planet had condensed to the liquid state before passing perihelion, which would certainly be true if its mass was less than 10^{28} grams (in other words, for all the terrestrial planets), it would be too dense to be broken up at all by tidal action. For such a planet could exist actually in contact with the sun's surface if its density was as great as fourteen times the mean density of the sun, which limit would be much exceeded by a liquid planet revolving about a giant sun.

This argument is not applicable to the great planets, for they might be still distended at their first perihelion passage, and might not reach a density that would forbid disruption until after a few revolutions.

It has been pointed out that the falling back into the sun of temporarily expelled matter would account for the sun's rotation; similarly the rotation of the planets may be explained. It is suggestive that the only planets with swift direct rotations are those whose direct-moving satellites indicate that they have been broken up by tidal action. The angular momenta of revolution of the satellites of the great planets are, however, much less than those of the rotations of the planets, so that this hypothesis cannot be accepted without a great deal of examination.

With regard to the method of condensation of the bodies produced, it is practically certain that the great planets liquefied by the process of 2·5 and the small satellites by that of 2·51. The method of solidification of the terrestrial planets and of the large satellites is not yet clear; it is a question of deciding at what mass the former process would become inoperative, which cannot be done with definiteness on account of the uncertainty in the temperature and density of the primitive filaments. The terrestrial planets, however, seem to be fundamentally different in constitution from the outer ones, suggesting that they lost much of their lighter constituents during condensation, and therefore that the critical mass was between those of Uranus and the earth, in accordance with the estimate of 2·4 (10); but the matter is not certain, and it is quite possible on present knowledge that the critical mass was between those of Titan and Iapetus.

2·9. *Summary.* A theory of the origin of the solar system, partly based on that of Jeans, has been developed. The sun, while in the giant stage, is supposed to have been broken up by the tidal action of a passing star several times more massive than itself. It is unlikely that a gaseous star of the sun's mass could have had a radius greater than 21 million km., which fixes an upper limit to the size of the sun at the time of the encounter. The encounter did not approximate either to Jeans's 'slow' or to his

* See later, Chap. IV.

'transitory' type. A slow encounter is dynamically impossible, while the matter ejected in a true transitory encounter would all fall back into the sun and be reabsorbed.

Accordingly the encounter must have been of intermediate type. The ejected matter, as it emerged from the sun, collected into nuclei. These continued to move outwards, but were deflected sideways by the star, and thus proceeded to move around the sun in one direction and nearly in one plane. If the diameter of the sun at the time was the greatest possible, the masses of the nuclei able to retain the whole of their constituents must have been greater than $1{\cdot}2 \times 10^{28}$ grams, about twice the mass of the earth. The great outer planets must therefore have retained all their constituents, whereas the smaller ones may have lost a large part of their mass; thus the fact that the great planets have low densities may be explicable.

The condensation of the great planets was a straightforward process, each passing steadily through the gaseous state to the liquid state through loss of heat by radiation from the outside. The smaller ones and the satellites had a more complicated history. They would form drops at the outside, and these would fall in towards the centre, forming liquid cores; but at the same time a large fraction of their mass would probably be lost.

When the terrestrial planets had reached the liquid state they might continue to cool adiabatically by evaporation from the surface; but evaporation alone would not have sufficed to bring them to the solid state, for it would have become inappreciable before they had cooled so far. Thus all the planets must have gone through a liquid stage, passing gradually into the solid state by cooling from the surface.

Most of the satellites were probably formed by the tidal disruption of their primaries by the sun when they passed perihelion for the first time. This hypothesis accounts readily for the general resemblance of the subsystems of the great planets to the solar system as a whole, and for the dissimilarity of the subsystems of the terrestrial planets. It suggests also that a number of satellites were detached completely from their original parents and became for a time independent planets. The moon cannot be explained in this way, as will be seen in Chapter III, where a different explanation of its origin will be given. The possible subsequent history of the lost satellites will be discussed in 4·4.

CHAPTER III

The Origin of the Moon

"There is a tide in the affairs of men
Which taken at the flood leads on to fortune."

SHAKESPEARE, *Julius Caesar*, IV, 3.

3·1. Our moon has probably had a very different origin and history from any other satellite. Although the earth is the second smallest planet to possess a satellite at all, it happens that the moon is the third or fourth most massive satellite in the whole system. Its mass is $\frac{1}{82}$ of that of the earth, while the best estimates available of the masses of other satellites indicate that the only heavier ones are J III, J IV, and Titan, whose masses are respectively $\frac{1}{40}$, $\frac{1}{70}$, and $\frac{1}{50}$ of that of the earth. The ratios of the masses of these large satellites to those of their primaries are only $\frac{1}{12500}$, $\frac{1}{22200}$, and $\frac{1}{4700}$ respectively. The remarkably large mass of the moon in proportion to its primary suggests two alternatives. It may have been formed from the earth in some way decidedly different from the births of the satellites of the great planets from their primaries; or it may have been formed from some other planet, but have left the latter immediately and become an independent planet. It will be seen later (4·4) that it is possible that some satellites were formerly independent planets, and have been captured by their present primaries; but in the particular case of the moon this possibility can be easily dismissed. To suppose it to have been formed from some other terrestrial planet would intensify the difficulty with regard to its mass. To suppose it to have been formed from an outer planet is practically impossible, for the following reason. It has been pointed out that the tidal theory strongly suggests that the more remote planets should have lower densities than the nearer ones. The same should be true of satellites, and it actually appears that the densities of the four great satellites of Jupiter decrease with distance from their primary. But a lost satellite would be originally more remote than any retained one, and should therefore have a lower density than any of the others. The moon, on the other hand, has a greater density than any other known satellite, nearly three times that of J III and four times that of J IV.

We must therefore retain the hypothesis that the moon was formed from the earth, although it cannot have been formed in the way already described. It is necessary that some other influence shall have cooperated in the fission. Now the density of the moon is such that if it were homogeneous it would be broken up by tidal action if it passed within 1·8 times the earth's radius from the centre of the earth; the moon must therefore have been formed at a distance from the centre of the earth greater than

this. Hence considerable extension of the primitive earth is necessary, although the solar tides can never unaided have been able to produce such an extension.

3·2. The most plausible theory of the origin of the moon is that suggested by Sir G. H. Darwin. Consider for a moment a man in a swing. If he is pulled back for a moment and released, he performs a few oscillations and gradually comes to rest. But if we give him a push every time he is nearest to us, the extent of the oscillations becomes greater every time, and if friction were absent, it would be possible to increase the amplitude indefinitely. This increase in the extent of a vibration when the external force has a period equal to the natural period is known as 'resonance.' If the earth, when it was wholly or partially liquid, received a small distortion so that its equator became an ellipse, it would oscillate backwards and forwards about the symmetrical form until friction brought it to rest. The period of this motion would be about two hours. But we can show that the angular momentum of the earth-moon system is such that, if the earth and moon ever formed a single body, this must have rotated in about four hours. Now the solar semidiurnal tide is just such an oscillation as we have been considering, and the period of the disturbance in height at any station is half the period of rotation, in this case two hours. Hence the amplitude of the oscillation would grow to be very great. There appears to be no limit to the amplitude of the tide that could be produced in this way. For suppose the oscillation to have continued long enough for its amplitude to have become steady, and consider the elevation of the surface at the point vertically below the sun. If the free period was shorter than the period of the tide, the tidal elevation would be positive, in accordance with the ordinary theory of forced oscillations; but if the period of the tide was the shorter, the elevation would be negative. Hence if the earth was condensing slowly, both periods changing slowly, and the one that was formerly the shorter became the longer, the elevation would change its sign. It could do this only by passing through zero or infinity. The former alternative implies that when perfect resonance is attained there would be no tide, which is unplausible; the latter implies that if only the conditions for resonance persisted for a long enough time there is no limit to the extent of the tide that could be produced. Thus the earth would be gradually stretched out in the direction of the sun. When the disturbance became great enough, the mass would break into two parts in much the same way as Jeans showed to be possible for a homogeneous liquid mass undergoing a slow tidal encounter without resonance.

Love* and Bryan† determined the angular velocity that the earth would have had to have in order to produce a satellite in this way, sup-

* *Phil. Mag.* (5), **27**, 1889, 254–64. † *Phil. Trans.* **153 A**, 1889, 187–219.

posing the earth to have been homogeneous, but Moulton* showed that the angular momentum in the system is too small to give the velocity of rotation required. It was found by the present writer, however, that when we take into account the fact that the earth is not homogeneous, the conditions become much more favourable to the theory. The oscillations of a heterogeneous mass in three dimensions have so far presented an intractable problem, but the corresponding two-dimensional problem has been solved, and it is found that homogeneity is the least favourable case to the theory that is possible. The modification produced in the numerical results by allowing as much heterogeneity in the two-dimensional problem as exists in the actual earth is found to be as great as is required to remove the discrepancy found by Moulton.

3·21. *The Free Vibrations of a Heterogeneous Liquid Cylinder.* In the equilibrium state the liquid is supposed to form two layers, an inner circular cylinder of density $\rho\,(1+\eta)$ and radius $a_1 = a\alpha$, surrounded by an outer layer of density ρ and radius a. Let the axis of z be along the centre of the cylinder, and suppose a constraint to prevent motion along the cylinder. Let the angular velocity be ω. The equations of motion, referred to rectangular axes rotating with this speed, are

$$\left.\begin{aligned} \dot{u} - 2\omega v - \omega^2 x &= \frac{\partial}{\partial x}\left(U - \frac{p}{\rho}\right) \\ \dot{v} + 2\omega u - \omega^2 y &= \frac{\partial}{\partial y}\left(U - \frac{p}{\rho}\right) \end{aligned}\right\} \qquad \text{.........(1)},$$

where u and v are assumed small and U is the gravitation potential; the equation of continuity is

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \qquad \text{.........(2)}.$$

By cross-differentiation we at once find

$$\frac{\partial}{\partial t}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) = 0,$$

so that in all periodic motions

$$\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 0 \qquad \text{.........(3)}.$$

Hence there are a velocity potential Φ and a stream function Ψ giving the motion relative to the rotating axes. Then

$$\nabla^2\Phi = 0 \quad \text{and} \quad \nabla^2\Psi = 0 \qquad \text{.........(4)}.$$

This will hold for both layers.

Let the equations of the bounding layers in cylindrical polar coordinates in the disturbed position be respectively

$$\left.\begin{aligned} r &= a + q\cos(n\phi + \beta) = a + qS, \quad \text{say} \\ r &= a_1 + q_1\cos(n\phi + \gamma) = a_1 + q_1 S_1, \quad \text{say} \end{aligned}\right\} \qquad \text{.........(5)},$$

where β and γ may be functions of the time and q and q_1 are small.

* T. C. Chamberlin and others, *The tidal and other problems*, Carnegie Institute.

Then the gravitation potentials in the outer and inner layers are

$$U_1 = -\pi f\rho(r^2-a^2) - 2\pi f\rho\eta a_1^2 \log\frac{r}{a_1} + \frac{2\pi f\rho}{n}\left\{aqS\left(\frac{r}{a}\right)^n + \eta a_1 q_1 S_1\left(\frac{a_1}{r}\right)^n\right\} \quad \ldots(6),$$

$$U_2 = -\pi f\rho(r^2-a^2) - \pi f\rho\eta(r^2-a_1^2) + \frac{2\pi f\rho}{n}\left\{aqS\left(\frac{r}{a}\right)^n + \eta a_1 q_1 S_1\left(\frac{r}{a_1}\right)^n\right\} \quad \ldots(7).$$

When $r = a$, let the part of $-\dfrac{\partial U}{\partial r}$ that is independent of ϕ be g; and when $r = a_1$, let the part independent of ϕ be g_1.

We have further

$$U - \frac{p}{\rho} = \int (\dot{u} - 2\omega v - \omega^2 x)\,dx + (\dot{v} + 2\omega u - \omega^2 y)\,dy$$
$$= \frac{\partial\Phi}{\partial t} - \tfrac{1}{2}\omega^2 r^2 + 2\omega\Psi \quad \ldots\ldots\ldots(8).$$

Assume next that in the inner and outer layers the velocity potentials have the forms

$$\Phi_1 = \left(\frac{r}{a}\right)^n K_1 + \left(\frac{a}{r}\right)^n K_2 \quad \ldots\ldots\ldots(9),$$

$$\Phi_2 = \left(\frac{r}{a_1}\right)^n K_3 \quad \ldots\ldots\ldots(10),$$

where K_1, K_2, and K_3 are harmonic functions of $n\phi$. Then the stream functions are obtained by writing

$$n\phi - \frac{\pi}{2} \text{ for } n\phi \text{ in } K_1 \text{ and } K_3, \text{ and } n\phi + \frac{\pi}{2} \text{ for } n\phi \text{ in } K_2 \ldots\ldots(11).$$

Then the boundary conditions give

$$n(K_1 - K_2) = a\frac{d}{dt}(qS) \quad \ldots\ldots\ldots(12),$$

$$n(K_1\alpha^n - K_2\alpha^{-n}) = a\alpha\frac{d}{dt}(q_1S_1) = nK_3 \quad \ldots\ldots\ldots(13).$$

Hence

$$K_3 = \frac{a\alpha}{n}\frac{d}{dt}q_1S_1 \quad \ldots\ldots\ldots(14),$$

$$K_1 = \frac{d}{dt}\frac{a(qS - q_1S_1\alpha^{n+1})}{n(1-\alpha^{2n})};\quad K_2 = \frac{d}{dt}\frac{a(qS\alpha^{2n} - q_1S_1\alpha^{n+1})}{n(1-\alpha^{2n})} \quad \ldots(15).$$

At the outer surface $\Psi = \dfrac{d}{dt}\dfrac{aq}{n}S\left(n\phi - \dfrac{\pi}{2}\right)$;

at the inner surface $\Psi = \dfrac{d}{dt}\dfrac{a\alpha}{n}q_1S_1\left(n\phi - \dfrac{\pi}{2}\right)$.

At the outer surface the pressure is constant. Therefore

$$U - \frac{\partial\Phi}{\partial t} + \omega^2 aqS - 2\omega\Psi = \text{constant},$$

or

$$-gqS + \frac{2\pi f\rho a}{n}(qS + \eta\alpha^{n+1}q_1S_1) + \omega^2 aqS$$
$$-\frac{d^2}{dt^2}\frac{a\{qS(1+\alpha^{2n}) - 2\alpha^{n+1}q_1S_1\}}{n(1-\alpha^{2n})} - \frac{2\omega}{n}\frac{d}{dt}aqS\left(n\phi - \frac{\pi}{2}\right) = 0 \quad \ldots(16).$$

At the inner surface the pressure is continuous. Hence

$$\eta\left[-g_1 q_1 S_1+\frac{2\pi f\rho a}{n}(qSa^n+\eta q_1 S_1 a)+\omega^2 a a q_1 S_1-2\omega\frac{aa}{n}\frac{d}{dt}q_1 S_1\left(n\phi-\frac{\pi}{2}\right)\right]$$
$$-(1+\eta)\frac{aa}{n}\frac{d^2}{dt^2}(q_1 S_1)+\frac{d^2}{dt^2}\frac{a\{2qSa^n-q_1 S_1 a(1+a^{2n})\}}{n(1-a^{2n})}=0 \quad\text{......(17)}.$$

So far nothing has been assumed about the form of the wave, save that S and S_1 are circular functions of the nth order. Suppose now that the wave travels with its form unchanged and that the speed of the vibration at any point is p. Then q and q_1 are constants, and S and S_1 are of the form $\cos(n\phi-pt+\beta)$.

Then $\frac{d}{dt}S\left(n\phi-\frac{\pi}{2}\right)=-pS$, and $\frac{d}{dt}S_1\left(n\phi-\frac{\pi}{2}\right)=-pS_1$ ......(18).

Thus

$$qS\left[-\frac{ng}{a}+2\pi f\rho+n\omega^2+p^2\frac{1+a^{2n}}{1-a^{2n}}+2\omega p\right]$$
$$+q_1 S_1\left[2\pi f\rho\eta a^{n+1}-\frac{2p^2 a^{n+1}}{1-a^{2n}}\right]=0 \quad\text{.........(19)},$$

$$qS\left[2\pi f\rho\eta a^{n-1}-\frac{2p^2 a^{n-1}}{1-a^{2n}}\right]$$
$$+q_1 S_1\left[-\eta\frac{ng_1}{aa}+2\pi f\rho\eta^2+n\eta\omega^2+(1+\eta)p^2+p^2\frac{1+a^{2n}}{1-a^{2n}}+2\eta\omega p\right]=0$$
$$\text{.........(19}\,a\text{)}.$$

The period equation is therefore

$$\left(-\frac{ng}{a}+2\pi f\rho+n\omega^2+2\omega p+p^2\frac{1+a^{2n}}{1-a^{2n}}\right)$$
$$\left(-\frac{\eta ng_1}{aa}+2\pi f\rho\eta^2+\eta n\omega^2+2\eta\omega p+\eta p^2+\frac{2p^2}{1-a^{2n}}\right)-\left(\pi f\rho\eta-\frac{p^2}{1-a^{2n}}\right)^2 4a^{2n}=0$$
$$\text{.........(20)}.$$

In the homogeneous case $\eta=0$, and this equation reduces to

$$p^2+2\omega p+n\omega^2+2\pi f\rho-ng/a=0,\quad p^2=0 \quad\text{.........(21)}.$$

The zero roots give merely a displacement of the inner boundary without the outer being affected, which is obviously possible when the densities are equal.

In the other type of oscillation the sum of the two possible speeds is -2ω, as was proved by Bryan for the corresponding three-dimensional problem. Further, in the case of the form of bifurcation ($n=2$, $p=0$) $\omega^2=\pi f\rho$, as was found by Jeans*.

We wish to know whether heterogeneity causes resonance or instability to occur for a smaller value of the angular velocity when the mean density is the same. Now the mean density is $\rho(1+\eta a^2)$, and hence the angular velocity that would cause both these conditions in a homogeneous cylinder with the same mean density is given by

$$\omega_1^2=\pi f\rho(1+\eta a^2) \quad\text{.........(22)}.$$

* 'The Equilibrium of Rotating Liquid Cylinders,' *Phil. Trans.* **200 A**, 1903, 81.

Resonance will occur in the heterogeneous case if one value of p is -2ω. In this case the angular velocity ω_2 is given by

$$\frac{8\omega^2}{1-\alpha^{2n}}\left\{-\frac{ng}{a}-\frac{\eta n g_1}{a}\alpha^{2n-1}+2\pi f\rho\,(1+2\eta\alpha^{2n}+2\eta^2\alpha^{2n})+n\omega^2\,(1+\eta\alpha^{2n})\right\}$$

$$+\,\eta\left(n\omega^2+2\pi f\rho-\frac{ng}{a}\right)\left(n\omega^2+2\pi f\rho\eta-\frac{ng_1}{a\alpha}\right)-4\pi^2f^2\rho^2\eta^2\alpha^{2n}=0\ \ldots(23).$$

Instability will commence with an angular velocity ω_3 given by putting $p = 0$. Then

$$\left(n\omega^2+2\pi f\rho-\frac{ng}{a}\right)\left(n\omega^2+2\pi f\rho\eta-\frac{ng_1}{a\alpha}\right)-4\pi^2f^2\rho^2\eta\alpha^{2n}=0\ \ldots\ldots(24).$$

As we are restricted to a two-dimensional problem and only require approximate results to indicate the direction of the effect of heterogeneity, accurate correspondence with the actual case is unnecessary. We shall assume

$$\rho = 3{\cdot}2,\quad \eta = 1{\cdot}50,\quad \alpha = 0{\cdot}66.$$

Then $\rho\,(1+\eta) = 8{\cdot}0;\ g/a = 3{\cdot}30\pi f\rho;\ g_1/a\alpha = 5{\cdot}00\pi f\rho$.........(25).

This makes the two densities, and the ratio of the amounts of matter of the two densities, about the same as in Wiechert's hypothesis of the structure of the earth. Then

$$\omega_1{}^2/\pi f\rho = 1{\cdot}65 \quad ....(26).$$

Putting $n = 2$ we find

$$\omega_2{}^2/\pi f\rho = 0{\cdot}71 \text{ or } 2{\cdot}10 \quad ....(27),$$

$$\omega_3{}^2/\pi f\rho = 2{\cdot}03 \text{ or } 3{\cdot}74 \quad ....(28).$$

If a widely diffused cylindrical mass with the density distributed in this way condenses so as to keep α constant, ρ and ω increase like a^{-2}. Thus $\omega^2/\pi f\rho$ increases like a^{-2}, and resonance will occur when it reaches the value 0·71, which is only 0·43 of the value needed to produce resonance and instability in a homogeneous mass with the same mean density. Thus heterogeneity encourages resonance; the fact that the smaller value of $\omega_3{}^2$ is greater than $\omega_1{}^2$ indicates that it discourages instability.

In general, put $p = k\omega$ and $\omega^2/\pi f\rho = \lambda$. Then the period equation for $n = 2$ is

$$\left\{-(1+2\eta\alpha^2)+\lambda\left(1+k+\tfrac{1}{2}k^2\frac{1+\alpha^4}{1-\alpha^4}\right)\right\}$$

$$\left\{-(2+\eta)+\lambda\left(1+k+\tfrac{1}{2}k^2+\frac{k^2}{\eta\,(1-\alpha^4)}\right)\right\}\eta-\left\{\eta-\frac{k^2\lambda}{1-\alpha^4}\right\}^2\alpha^4=0\ \ \ldots(29).$$

This may be regarded as a quadratic in λ when k is known. The solution is given in the table. In the last two columns λ_0 is the value of $\omega^2/\pi f\rho$ for a homogeneous cylinder of density $\rho\,(1+\eta\alpha^2)$, so that a direct comparison is obtained between heterogeneous and homogeneous cylinders of the same mean density.

k	λ		$\lambda^{\frac{1}{2}}$		$k\lambda^{\frac{1}{2}}$		$\lambda_0^{\frac{1}{2}}$	$k\lambda_0^{\frac{1}{2}}$
∞	0	0	0	0	1·53	2 06	0·00	1·82
3	0·200	0·254	0·447	0·504	1·34	1·51	0·44	1·33
2	0·367	0·423	0·606	0·650	1·21	1·30	0·57	1·15
1	0·83	1·05	0·91	1 03	0·91	1·03	0·81	0·81
0·5	1·30	1·95	1·14	1·40	0·57	0 70	1·01	0·50
0	2·03	3·74	1·42	1·93	0·00	0·00	1·28	0·00
−0·5	2·83	5·43	1·68	2·33	−0·84	−1·12	1·65	−0·83
−1·0	2·34	4·23	1·53	2·06	−1·53	−2·06	1·82	−1·82
−2·0	0·71	2·10	0·84	1·45	−1·68	−2·90	1·28	−2·57
$-\infty$	0	0	0	0	−1·53	−2·06	0 00	−1 82

In the diagram the full curves show the variation of $k\lambda^{\frac{1}{2}}$ with $\lambda^{\frac{1}{2}}$, or for masses with the same densities, of p with ω. The two branches of the curve approach very closely near $k = 2$, but do not intersect. The dotted

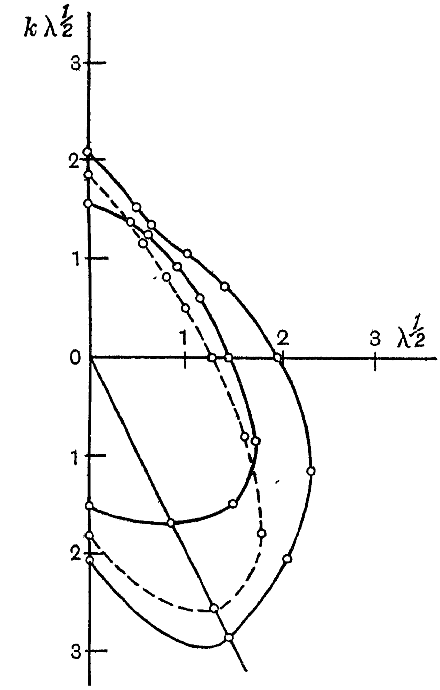


Fig. 3

curve is obtained by plotting $k\lambda_0^{\frac{1}{2}}$ against $\lambda_0^{\frac{1}{2}}$, so that it shows the speeds on the same scale for masses of the same mean density. The condition for resonance is given by the intersection of the curves with the line $p = -2\omega$ or $k = -2$, and it is at once seen that heterogeneity causes this to be satisfied for a much smaller value of the rotation than was otherwise needed. Further, if α be made to approach to unity, two of the values of k can be made to approach zero as closely as we like, while λ retains any finite value. Hence by making the depth of the outer layer small enough we can make resonance occur for as small a value of the

rotation as we please. It appears physically probable that the same will be true in the three-dimensional problem.

3·3. *The Original Period of Rotation.* Wiechert's hypothesis makes the principal moment of inertia of the earth

$$C = \tfrac{8}{15}\pi\rho a^5 (1 + \eta\alpha^5) \qquad \text{.........(1)},$$

and if M be the mass of the earth and m that of the moon

$$C = 0{\cdot}334 Ma^2 \qquad \text{.........(2)}.$$

Let ω and n be the respective angular velocities of the earth's rotation and the moon's revolution at present, and let c be the moon's distance. Then the angular momentum of the system is

$$C\omega + \frac{Mm}{M+m}c^2 n = 5{\cdot}78\, C\omega \qquad \text{.........(3)}.$$

Now $\omega = 7{\cdot}29 \times 10^{-5}/1$ sec. The moment of inertia of the combined body before separation would be approximately $C\,(M+m)^{\frac{5}{3}} m^{-\frac{5}{3}}$, apart from the increase due to flattening. Hence the angular velocity when the two formed one body was

$$5{\cdot}78\omega\left(\frac{M}{M+m}\right)^{\frac{5}{3}} = 4{\cdot}14 \times 10^{-4}/1 \text{ sec.},$$

provided that the density distribution was similar, and that the angular momentum and moment of inertia were not different. The only cause of change in angular momentum would be solar tidal friction, which probably would not amount to more than about 5 per cent. of the whole. Thus finally the angular momentum would be enough to make the whole rotate with an angular velocity of about $4{\cdot}3 \times 10^{-4}/1$ sec., apart from variations in the moment of inertia due to flattening and condensation.

Now a homogeneous ellipsoid would give similarly a maximum angular velocity of $3{\cdot}6 \times 10^{-4}/1$ sec., and instability in the symmetrical form and resonance would not occur until the rotation speed was $7{\cdot}2 \times 10^{-4}/1$ sec. Thus the homogeneous mass could never attain conditions suitable for resonance. The effect of heterogeneity on these conditions seems to be to increase the actual speed, by what has been said above, and to diminish the speed needed for resonance, so that the circumstances are much more favourable.

The problem of finding out the actual amount of the effect of heterogeneity on the period of rotation necessary to cause resonance is likely to be exceedingly difficult, as the bounding surfaces are not ellipsoids. An estimate can, however, be made by analogy with the cylindrical case, which has been accurately solved. In the two-dimensional case considered the angular velocity found necessary for resonance is only $0{\cdot}65$ of that needed when the mass is homogeneous. If the same ratio held in the three-dimensional case, the angular velocity needed would be $4{\cdot}7 \times 10^{-4}/1$ sec.,

while the available angular velocity is $4{\cdot}3 \times 10^{-4}/1$ sec. There are, however, two causes that will affect the former amount. Compressibility by introducing a further degree of freedom may be expected to reduce the free speed for any given angular velocity, and hence to reduce the angular velocity needed to make the ratio k equal to -2. Further, by reducing the relative thickness of the outer layer this free speed may be indefinitely reduced, and therefore it is evident that there can be a heterogeneous distribution of density, not very different from the actual distribution, such that with the same mean density the available angular momentum would lead to resonance. With still smaller relative thicknesses resonance would still be possible, but would occur at an earlier stage of condensation. The truth of the resonance theory is therefore highly probable.

3·4. *The Vibrations of a System with one degree of Freedom when the Free and Forced Periods are nearly equal and slowly varying.* Let the equation of motion of the system be

$$\ddot{x} + \mu\dot{x} + \alpha^2 x = Ee^{\iota pt} \qquad \text{.........(1)},$$

where μ is small, and α, p, and μ are slowly varying.

Put $\int \alpha dt = w$, and let accents denote differentiation with regard to w.

Then
$$x'' + \nu x' + x = E \exp \iota p \int \alpha^{-1} dw \qquad \text{.........(2)},$$

where
$$\nu = \frac{1}{\alpha}\frac{d\alpha}{dw} + \frac{\mu}{\alpha} \qquad \text{.........(3)}.$$

Now let one of the complementary functions of this be $\exp \theta$. Then it has been already shown* that the value of θ is practically $(\iota - \tfrac{1}{2}\nu)\, w$ for a considerable range in w.

Put $x = y \exp \theta$. Then

$$y'' + y'\,(2\theta' + \nu) = E \exp\left(\iota p \int \frac{dw}{\alpha} - \theta\right) \qquad \text{.........(4)}.$$

If p and α are nearly equal, we can write $p\int\frac{dw}{\alpha} = (1 + \beta w)\, w$, where β is small. Then the expression on the right of (4) is $E \exp\,(\tfrac{1}{2}\nu w + \iota\beta w^2)$, which varies slowly. Hence y'' can be neglected on the left, and the particular integral is approximately given by

$$y = \frac{E}{2\iota}\int \exp\,(\tfrac{1}{2}\nu w)\, dw = \frac{E}{\iota\nu} \exp\,(\tfrac{1}{2}\nu w)$$

near the instant when the periods coincide.

Hence $x = \dfrac{E}{\iota\nu} \exp\,(\iota w) + A \exp\,(\iota - \tfrac{1}{2}\nu)\, w + B \exp - (\iota + \tfrac{1}{2}\nu)\, w$......(5),

where A and B are arbitrary constants. The amplitude of the forced vibration is therefore magnified in the ratio $\alpha\Big/\left(\dfrac{d\alpha}{dw} + \mu\right)$ when the periods coincide.

* *Memoirs of R A S.* **60**, 1915, 211–13.

In the case of a fluid sphere of the same size as the earth, μ is small, and hence the magnification is

$$\alpha \Big/ \frac{d\alpha}{dw} \qquad \text{.........(6).}$$

Now the change of α in one period is $2\pi d\alpha/dw$, and hence the magnification is $2\pi\alpha \div$ the change of α in one period.

In the present case the equilibrium amplitude is about 50 cms., and the amplitude needed to cause rupture is presumably of order 10^8 cms. The condition for this is that the change of α in one period must have been less than $\pi\alpha \times 10^{-6}$, so that if α were varying so slowly as not to change greatly in a century, an amplitude of the requisite order of magnitude would be reached when the periods became equal. This hypothesis does not seem unreasonable; it may be considered then that there was time for resonance to produce the required amplitude.

3·5. The distorted form of the earth would be very long and narrow. The dense interior, instead of being drawn towards the sun, would be depressed, in accordance with the results of 3·21. Thus when the elongation became so great that the mass became unstable and the end broke off, the detached portion would be at a considerable distance from the centre of the earth, certainly several times the undisturbed radius, and it would be composed chiefly of materials from the outer regions of the earth. Its linear dimensions would be decidedly less than the radius of the earth, but of the same order of magnitude. The last two results agree with the actual size and density of the moon; it will be seen later (Chapter XIV) that the first has had an important influence on its history.

It should be noticed that, although the elongation would be towards the sun throughout the changes, the matter of the earth would always be rotating within the slowly moving surface, just as water can revolve within a fixed elliptical dish, each particle within the earth revolving about the axis in the period of rotation, whereas the surface would only complete its revolution in a year. Thus when rupture occurred the detached portion would have a considerable transverse velocity and therefore would not fall back into the earth.

CHAPTER IV

The Resisting Medium

"Friction produces heat." *Any School Physics.*

4·1. *Origin of the Medium.* So far we have seen that

(1) the disruption of the primitive sun by a passing star could have led to the formation of the planets;

(2) the tidal action of the sun on the outer planets, the first time they passed perihelion, may have led to the formation of satellites and of independent small planets;

(3) such planets as stayed with the sun, and such satellites as remained with their original primaries, would have had direct revolutions;

(4) the fact that the smallest planets are the densest is explicable on the same hypothesis;

(5) the fact that the planes of the motions of all these bodies are nearly coincident is similarly explicable;

(6) the moon may have been produced from the earth by the solar tides, magnified by resonance.

Several striking facts about our system, however, still remain unexplained. It is necessary to provide explanations of

(1) the smallness of the eccentricities of the orbits of the planets and satellites;

(2) the retrograde motions of two satellites of Jupiter, one of Saturn, four of Uranus, and one of Neptune;

(3) the curious numerical relations between the mean motions of several satellites;

(4) the retrograde rotations of Uranus and Neptune, and the direct rotations of the earth and Mars;

(5) the formation of asteroids;

(6) the acquirement by Mars of two small satellites;

(7) the recession to its present distance of the moon, which can have been only 10,000 to 20,000 km. from the centre of the earth when it was formed.

Some of these questions have been answered, and hitherto none has been proved unanswerable; but until all have been answered the theory of the origin and development of the solar system cannot be considered complete. The successes so far attained, however, are enough to encourage the cosmogonist to hope for the attainment of the others.

So far little explicit use of any form of friction has been made in the

theory, although it has been virtually assumed in the supposition that any ejected body reabsorbed into its parent would become an integral part of it, the whole rotating as a rigid body. Friction must, however, have influenced the history of the solar system in at least two other ways, namely, by the action of a resisting medium and by tidal friction, of which the former has probably had the more far-reaching effects, although it happens that tidal friction has been exceptionally important in determining the evolution of the earth and moon in particular (see later, Chapter XIV).

It has already been indicated that the matter ejected from the sun would not all be included in the planets and their satellites. Much of it would be lost on account of the inadequacy of the gravitative power of the distended nuclei to retain their lighter constituents, and the thinner parts of the filament would probably be unable to condense at all, but would spread out at once. This lost matter would be dispersed throughout the system, and would form the resisting medium. It is clear from its formation that it would be largely or entirely gaseous. It would have the same origin as the planets, and therefore would have been deflected transversely by the star, just as the planets were. Hence every part of it would have a direct revolution about the sun from the very beginning. The parts would, on the other hand, revolve in widely different periods, and would undergo diffusion at the same time, until the whole system was filled with tenuous matter. Differences in the periods of revolution would make some parts move outwards while others meeting them were moving inwards, and thus the radial motions would be quickly annulled by turbulence and viscosity. Thus the resisting medium would be a gas, its parts revolving around the sun in the same direction as the planets, and describing approximately circular paths. We have no knowledge of the composition of the matter originally ejected that might enable us to estimate the mass of the medium or the distribution of density within it. This can be found, if at all, only from the effects that we assume it to have produced. It will be found, however, that it leads to another inference that is capable of independent test.

4·2. *Density Distribution and Motion of the Medium.* Let us now consider the nature of the motion of the medium. For the reasons already given, the medium will be supposed to be symmetrical about an axis, which will be taken to be the axis of z, and every part of it will be supposed to revolve uniformly in a circle about this axis. Taking rectangular coordinates x and y in fixed directions in the equatorial plane of the mass and through the centre of the sun, we put

$$x^2 + y^2 = \varpi^2 \qquad \text{.........(1),}$$

and denote the velocity at any point by $\omega\varpi$ perpendicular to the meridian

plane. So far ω is unspecified and need not be a constant. Then the equations of motion of the medium are

$$\left.\begin{aligned}\frac{\partial U}{\partial x}-\frac{1}{\rho}\frac{\partial p}{\partial x}&=-\omega^2 x\\ \frac{\partial U}{\partial y}-\frac{1}{\rho}\frac{\partial p}{\partial y}&=-\omega^2 y\\ \frac{\partial U}{\partial z}-\frac{1}{\rho}\frac{\partial p}{\partial z}&=0\end{aligned}\right\}\qquad\text{.........(2),}$$

where U is the gravitation potential, p the pressure, and ρ the density. In differential notation these may be written

$$dU-\frac{1}{\rho}\,dp=-\tfrac{1}{2}\omega^2 d\varpi^2\qquad\text{.........(3).}$$

If V denote the absolute temperature, we have

$$p=R\rho V\qquad\text{.........(4),}$$

where R is a constant if the composition of the material is uniform. Also

$$U=fM/r\qquad\text{.........(5),}$$

where f is the constant of gravitation, M the mass of the sun, and

$$r^2=x^2+y^2+z^2\qquad\text{.........(6).}$$

The mass of the medium being supposed a small fraction of that of the sun, its gravitation may be neglected.

The temperature in an approximately steady state is determined by the condition that each part of the mass receives just as much heat as it radiates. The other parts of the envelope are by hypothesis much colder than the sun, and we may assume that each part is warmed wholly by solar radiation, and radiates its heat away at a rate proportional to V^4. The rate of receipt of heat is proportional to $1/r^2$, and therefore

$$V=ar^{-\frac{1}{2}}\qquad\text{.........(7),}$$

where a is some constant. Then equation (3) gives, after substitution from (4), (5) and (7),

$$\left(-\frac{fM}{ar^{\frac{3}{2}}}+\frac{1}{2}\frac{R}{r}\right)dr-R\frac{d\rho}{\rho}=-\frac{1}{2}\frac{\omega^2 r^{\frac{1}{2}}}{a}d\varpi^2\qquad\text{.........(8).}$$

It follows that $\omega^2 r^{\frac{1}{2}} d\varpi^2$ is a perfect differential, and

$$\omega^2=r^{-\frac{1}{2}}F^2\qquad\text{.........(9),}$$

where F is a function of ϖ, as yet unspecified. Substituting in (8) and integrating, we have

$$R\log\rho=\frac{2fM}{ar^{\frac{1}{2}}}+\tfrac{1}{2}R\log r+\frac{1}{2a}\int F^2 d\varpi^2+\text{const.}\qquad\text{......(10).}$$

Accordingly, since the integrand in the last variable term on the right is essentially positive, the density when r is great enough must increase with r at least as fast as $r^{\frac{1}{2}}$, and therefore the whole mass must be infinite. Hence

an exact steady motion is impossible with the conditions specified; the fluid must necessarily flow outwards to some extent. The loss will, however, not affect the distribution of density appreciably unless $fM/aRr^{\frac{1}{2}}$ is small. Supposing, as is reasonable, that the temperature 10^{13} cm. from the sun (roughly the distance of Venus) was 300° abs., (7) shows that a is 10^9 C.G.S. Cent. units, and if the medium be supposed to consist of hydrogen, R is 4×10^7 C.G.S. Then the ratio in question becomes equal to unity when r is 4×10^{19} cm. Thus outward flow would not be important in masses of the size of the solar system, and in them the second term on the right of (10) and the term in R/r in (8) are unimportant.

Suppose now that on the equatorial plane of the medium

$$\omega^2 = \lambda^2 f M \varpi^{-3} \qquad \text{.........(11)},$$

so that the velocity at any point is λ times what it would be if every part of the medium were describing a circle freely under gravity. Then (8) gives on the equatorial plane

$$R\frac{d\rho}{\rho} = \frac{fM}{ar^{\frac{3}{2}}}(\lambda^2 - 1)\,dr \qquad \text{.........(12)}.$$

Thus if the angular velocity at any point exceeds the circular velocity, the density will increase outwards, while if it is less than the circular velocity, the density will increase inwards. The more closely the velocity approximates to the circular velocity, the more nearly will the density be uniform. To indicate the importance of this approximation, let us consider the special case of no rotation, supposing the density at the edge of the primitive sun, where

$$r = 2 \times 10^{12}\ \text{cm.} \qquad \text{.........(13)},$$

to be equal to the maximum possible value, namely the density of the primitive sun, 5×10^{-5} gm./cm.^{3} Then with the data already adopted, the density at any point is given by

$$\log \frac{\rho}{5 \times 10^{-5}} = -\frac{1{\cdot}2 \times 10^{10}}{r^{\frac{1}{2}}} \qquad \text{.........(14)},$$

and the density near the orbit of Mars is therefore less than 10^{-1000}. Beyond the orbit of Mars it would be still less, and therefore the mass of the medium between the orbits of Mars and Neptune, the radius of Neptune's orbit being 5×10^{14} cm., is less than $\frac{4}{3}\pi . 10^{-1000}(5 \times 10^{14})^3$ grams, a quite inappreciable fraction of a gram, and totally incapable of ever producing a noticeable influence on the orbit of the smallest asteroid. Hence the resisting medium could not be of any cosmogonical importance unless each part of it revolved with very nearly the velocity appropriate to a planet moving in a circular orbit at the same distance. With very slight departures from this relation an extremely wide range of variation of density within the medium could be realized.

The last point requires considerable emphasis, since most writers that

have made use of a resisting medium in cosmogonical theories have assumed without investigation that such a medium would be at rest. A medium at rest would oppose a steady resistance to the motion of a planet, and would thereby reduce its total energy and make it fall towards the sun. Thus a stationary resisting medium would cause all planets to approach the sun, and this result has been habitually assumed in discussions of the effects of such a medium. From what has just been shown it appears that the medium would move in such a way that a planet in a circular orbit would have no motion relative to it, and therefore would experience no resistance and be quite unaffected by it.

4·3. *Effect of a Planet on the Medium.* In the case of a small satellite moving in a variational orbit around a planet and the sun, the corresponding proposition would be that the satellite would still always be moving in the same way as the medium around it; so that every part of the medium would have to move in a variational orbit, fluid pressure not affecting its motion in the least. This will be shown to be inconsistent with the equation of continuity, unless the temperature is the absolute zero.

4·31. It appears that the motion of the medium around a planet and the sun, the planet's orbit being circular, could not be even roughly steady. It can be shown that the only possible steady motions of a gaseous medium or a swarm of meteors are such that the whole medium is statistically rotating with the planet like a rigid body. For, let us take the axes of x and y through the sun, the axis of x being always towards the planet. If the mean motion of the planet be n and the gravitation potential U, the equations of motion of any particle are

$$\left.\begin{aligned} \frac{d^2x}{dt^2} - 2n\frac{dy}{dt} - n^2x &= \frac{\partial U}{\partial x} \\ \frac{d^2y}{dt^2} + 2n\frac{dx}{dt} - n^2y &= \frac{\partial U}{\partial y} \\ \frac{d^2z}{dt^2} \qquad\qquad\qquad &= \frac{\partial U}{\partial z} \end{aligned}\right\} \quad \ldots\ldots\ldots(1).$$

Multiplying by $\dfrac{dx}{dt}, \dfrac{dy}{dt}, \dfrac{dz}{dt}$ and adding, we obtain on integration the equation

$$\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2 - n^2(x^2+y^2) = 2U + C \quad \ldots\ldots\ldots(2),$$

where C is a constant throughout the motion of the particle. This is Jacobi's integral, and it is known that the equations of motion possess no other first integral.

Now consider a swarm of particles, which may be solid bodies or the molecules of a gas, revolving about the sun and planet. Let the number in the element of volume $dx\,dy\,dz$, the rates of change of whose coordinates

are between u and $u + du$, v and $v + dv$, and w and $w + dw$ respectively, be $\phi\,(uvwxyz)\,du\,dv\,dw\,dx\,dy\,dz$. (It is to be noticed that u, v, w are to be regarded as rates of change of the coordinates and not as velocity components referred to fixed axes momentarily coincident with the moving ones.) In the absence of collisions each particle will move in time dt to a position $x'y'z'$, and will be changing its coordinates at rates u', v', w', where to the first order in dt

$$x' = x + u\,dt, \text{ etc.,} \qquad u' = u + \frac{du}{dt}\,dt, \text{ etc.} \qquad \text{.........(3).}$$

But $\frac{du}{dt}$, etc., are known functions of the coordinates and their rates of change, so that x', y', z', u', v', w' are known as functions of x, y, z, u, v, w and of dt. All particles in the original element move in time dt into the new element, and therefore

$$\phi'\,(u'v'w'x'y'z')\,du'\,dv'\,dw'\,dx'\,dy'\,dz' = \phi\,(uvwxyz)\,du\,dv\,dw\,dx\,dy\,dz \;\text{...(4),}$$

where ϕ' indicates that the velocity and density distribution after time dt is being considered. If the distribution is to be steady, so that the number of particles in a given element as regards position and velocity will be the same for all time, ϕ' is the same as ϕ. Now

$$\begin{aligned}\frac{du'\,dv'\,dw'\,dx'\,dy'\,dz'}{du\,dv\,dw\,dx\,dy\,dz} &= \frac{\partial\,(u'v'w'x'y'z')}{\partial\,(uvwxyz)} \\ &= \begin{vmatrix} 1 & 0 & 0 & \frac{\partial}{\partial x}\frac{du}{dt}dt & \frac{\partial}{\partial y}\frac{du}{dt}dt & \frac{\partial}{\partial z}\frac{du}{dt}dt \\ 0 & 1 & 0 & \frac{\partial}{\partial x}\frac{dv}{dt}dt & \frac{\partial}{\partial y}\frac{dv}{dt}dt & \frac{\partial}{\partial z}\frac{dv}{dt}dt \\ 0 & 0 & 1 & \frac{\partial}{\partial x}\frac{dw}{dt}dt & \frac{\partial}{\partial y}\frac{dw}{dt}dt & \frac{\partial}{\partial z}\frac{dw}{dt}dt \\ dt & 0 & 0 & 1 & 0 & 0 \\ 0 & dt & 0 & 0 & 1 & 0 \\ 0 & 0 & dt & 0 & 0 & 1 \end{vmatrix} \\ &= 1 + \text{terms in } dt^2 \qquad \text{.........(5).}\end{aligned}$$

It is proved in works on the dynamical theory of gases that collisions do not affect this result (cf. Jeans, *Dynamical Theory of Gases*, 2nd edition, p. 226).

Substituting in (4) we see that

$$\phi\,(u'v'w'x'y'z') = \phi\,(uvwxyz) \qquad \text{.........(6)}$$

to the first order in dt; so that

$$\frac{d\phi}{dt} = 0 \qquad \text{.........(7),}$$

if u', v', w', x', y', z' are related to u, v, w, x, y, z according to the relations (3), with the values of $\frac{du}{dt}$, $\frac{dv}{dt}$, $\frac{dw}{dt}$ given by (1). In other words $\phi =$ constant is a first integral of the equations of motion. Further, since by hypothesis the

motion is steady, ϕ does not involve the time explicitly. Thus this theorem, proved by Jeans for fixed axes, is readily extended to moving axes.

In the case we are considering there is no first integral except the Jacobi integral. It follows that in a steady state ϕ must be a function of

$$u^2 + v^2 + w^2 - n^2(x^2 + y^2) - 2U = u^2 + v^2 + w^2 - \nu, \text{ say} \quad \ldots(8).$$

Put then

$$\phi = \chi(\nu - u^2 - v^2 - w^2) \quad \ldots\ldots\ldots(9).$$

The numerical density λ can be found by putting

$$u = \xi\nu^{\frac{1}{2}}, \quad v = \eta\nu^{\frac{1}{2}}, \quad w = \zeta\nu^{\frac{1}{2}} \quad \ldots\ldots\ldots(10),$$

when

$$\begin{aligned}\lambda &= \iiint \phi \, du \, dv \, dw \\ &= \iiint \chi\{\nu(1 - \xi^2 - \eta^2 - \zeta^2)\} \nu^{\frac{3}{2}} d\xi \, d\eta \, d\zeta \quad \ldots\ldots\ldots(11),\end{aligned}$$

the limits being $-\infty$ to ∞ in all cases. Hence λ, and therefore ρ, are functions of ν alone. Since ϕ is an even function of u, v, w, we see further that the medium as a whole has no systematic motion with reference to the axes. Both of these results are independent of whether the medium consists of a gas or of solid particles.

4·32. The first of these results finds a simple application in connection with the Moulton-Gyldèn theory of the Counterglow*. According to this theory the counterglow is caused by light reflected from particles describing orbits about the sun and earth jointly, a particularly large number of which are visible at any time in the part of the sky directly opposite to the sun. From the result that the density is a function of ν alone we infer that the illumination will be greatest where the observer is looking through the greatest depth between consecutive surfaces of the system $\nu =$ constant. But just opposite to the sun is a place where one of these surfaces has a conical point, and it is at this point that the distance between surfaces of the system is greatest, just as a hyperbola is furthest from its asymptotes near the centre. Hence at this point the observer is looking through the greatest depth of matter, and therefore sees a patch of reflected light.

4·33. Next, consider the motion of a small particle through the medium. The resistance to its motion will be opposite to the direction of its velocity relative to the medium, and will vanish with the relative velocity. But the medium has no systematic motion with regard to the axes. If then the coordinates of the particle are x, y, z, the components of the retardation will be $-\kappa\dot{x}$, $-\kappa\dot{y}$, $-\kappa\dot{z}$, where κ is positive. Hence its equations of motion are

$$\left.\begin{aligned}\frac{d^2x}{dt^2} - 2n\frac{dy}{dt} - n^2x &= \frac{\partial U}{\partial x} - \kappa\frac{dx}{dt} \\ \frac{d^2y}{dt^2} + 2n\frac{dx}{dt} - n^2y &= \frac{\partial U}{\partial y} - \kappa\frac{dy}{dt} \\ \frac{d^2z}{dt^2} \qquad\qquad\qquad &= \frac{\partial U}{\partial z} - \kappa\frac{dz}{dt}\end{aligned}\right\} \quad \ldots\ldots\ldots(1).$$

* *Bulletin Astronomique*, t. 1; *Astronomical Journal*, No. 483.

Multiply by $\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}$ respectively and add. Then

$$\frac{d}{dt}\left\{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2+\left(\frac{dz}{dt}\right)^2-n^2(x^2+y^2)-2U\right\}=-2\kappa\left\{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2+\left(\frac{dz}{dt}\right)^2\right\} \qquad \text{.........(2).}$$

Hence the function $\nu - \dot{x}^2 - \dot{y}^2 - \dot{z}^2$ steadily increases with the time. Thus if we denote it by μ, we see that ν is always greater than μ, which is itself in general steadily increasing at a finite rate. Hence in time μ, and therefore ν, will exceed any finite limit. Whatever be the value of μ at any one time, the motion will be such that at any subsequent time the particle will be at a place where ν is greater than that value of μ. Now the surfaces $\nu =$ constant are closed, and the greater ν is the more closely they approach the three places where ν becomes infinite, situated at the sun, the planet, and at an infinite distance. Thus the motion of the third body will become more and more restricted until it is ultimately forced to revolve as an independent planet, as a satellite of the planet, or is expelled from the system; in either of the two former events it will steadily approach its primary. The only exceptional case is that where the coordinates of the third body remain the same permanently, so that $\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2+\left(\frac{dz}{dt}\right)^2$ is zero. In such a case friction will not affect the motion. The only stable motion of this type is that where the three bodies are at the corners of an equilateral triangle, and in this case ν is a minimum. Hence if any small displacement from the equilateral triangle position takes place, the disturbance will steadily increase until the third body becomes either an independent planet or a satellite. We notice also that a small body with a greater mean distance than the planet must approach the sun and hence become either an interior planet or a satellite. It thus appears possible that the capture of satellites can occur if the resisting medium is in a state of steady motion.

The above argument, however, cannot be applied entirely to the actual evolution of the solar system. In a gaseous medium ϕ must be of the form $Ne^{h^2(\nu-u^2-v^2-w^2)}$, where N and h are constants. It follows that the density is proportional to $e^{h^2\nu}$. It can hence be easily shown that the results based on 4·2 (12) are not appreciably altered in the regions remote from the planet. In the neighbourhood of Jupiter, again, the variation of the temperature would not be considerable, and we should have nearly

$$\log \rho = \frac{\nu}{RV}.$$

Taking $V = 100°$, we see that the difference between the values of $\log \rho$ at the surface of Jupiter and 100 radii away is of the order of 5000, indicating, as in 4·2, that the assumption that the more remote satellites of Jupiter have been appreciably affected by the resisting medium is inconsistent with the assumptions of a reasonable density near Jupiter and

of steady motion of the medium. It appears, therefore, that the medium cannot have been even locally in a state of steady motion, and that the results obtained on this hypothesis, though of interest as suggesting possible modes of development, cannot be regarded as giving the correct theory of the influence of a resisting medium on satellites.

4·34. We may, however, recur to the hypothesis of 4·3 that the medium at any point was moving like a planet revolving in a periodic orbit through the point. This velocity is fixed for given coordinates, and therefore the motion would be a steady one: but for any other temperature than the absolute zero we have found the only steady motion dynamically possible, and it is not this one. Thus a steady motion of this type is possible only if the temperature is the absolute zero. But the temperature cannot be the absolute zero. Hence this type of motion is impossible, and the medium must exert a dissipative influence on the third body. Some secular effect on the mean distance of a satellite from its primary is therefore probable, though not certain.

4·4. *Effect of the Resisting Medium on the Mean Distances of Satellites.* Now considering the solar system as it is, we notice several striking facts that are readily explicable if it is true that a resisting medium would make satellites approach their primaries, and might even enable a planet to capture a body previously moving as an independent planet and force it to move as a satellite of its own. The retrograde satellites of Jupiter and Saturn are readily explained on the hypothesis that they were produced by the tidal action of the sun on the planets, but left their parents at once and became independent planets; and that they afterwards were captured by their present primaries, which may or may not have been their original parents. The same may also apply to some direct satellites, especially to the two satellites of Mars, the sixth and seventh of Jupiter, and Iapetus. Such an abnormal origin is suggested in the case of the satellites of Mars by their small size; such small bodies could not have been formed from a planet such as Mars, which must have liquefied almost immediately. For J VI, J VII, and Iapetus it is suggested by the fact that whereas the inner satellites of Jupiter and Saturn are regularly spaced as regards distance from the primary, wide gaps separate them from the orbits of these outer ones. It is possible, however, that the gap is due to the density of the medium at the distance of these satellites having been very small in comparison with that nearer the primaries, so that the inner satellites were led to approach their primaries, leaving the outer ones almost unaffected.

4·41. The hypothesis that a resisting medium in some cases did make satellites approach their primaries is strongly confirmed by the existence of Saturn's rings. It is generally believed that the rings represent the fragments of a solid satellite that was broken up by the tidal action of

Saturn through being too near to the primary for the mutual attractions of its parts to resist the disruptive tendency. Now if the rings were initially at their present distance, they could never have condensed; for the efficacy of the tidal action in comparison with the mutual gravitation of the parts would increase with distension, and therefore if the former were able to disrupt it when solid, still more capable would it be when gaseous. Hence the mass would have remained gaseous until it diffused away. Hence the satellite that formed the ring must have been formed when beyond the danger zone and have afterwards been brought within it by some disturbing influence. No other agency than a resisting medium has been suggested that could produce such an approach.

4·42. Several curious numerical relations hold between the mean motions of various satellites. The mean motions n_1, n_2, n_3 of J I, J II, and J III are approximately in the ratios 4 : 2 : 1, while $n_1 - 2n = n_2 - 2n_3$ are exactly equal to each other and approximately to $\frac{1}{300}n_1$.

The mean motions of Mimas, Enceladus, Tethys, and Dione are nearly in the ratios 6 : 4 : 3 : 2, none departing from the values corresponding to these ratios by more than 3 per cent. Those of Titan and Hyperion are as 4·004 : 3. On any theory yet advanced of the origin of satellites it is very difficult to see how such relations could have subsisted from the beginning; but if the mean distances of the satellites were varying continually owing to a resisting medium, such ratios would occur several times during the evolution, and if the corresponding states of the systems were stable they would thenceforth persist.

4·43. The satellites of Uranus and Neptune were probably formed by tidal disruption of their primaries by the sun. Any satellite formed in this way would be direct; but if the axis of rotation of the primary was strongly inclined to the ecliptic, the ellipticity of figure of the primary would make the plane of the satellite's orbit revolve; and if a resisting medium were available to damp down the component of the motion of the satellite parallel to the axis of the planet, the satellite would come to revolve in the plane of the equator of the primary, even though the primary might have a retrograde rotation.

4·5. *Evolution of the Medium, and its Effect on Mercury.* Let us now consider the manner of evolution of the resisting medium. So far its internal viscosity, diffusion, and thermal conductivity have all been ignored. Since viscosity must necessarily produce a secular effect on the motion of any mass that is not moving either irrotationally or with the same rotation at all points, neither of which conditions is satisfied by a medium moving as this one would, the motion of the medium must undergo a slow and steady change. The nature of the change is easily seen. The fast-moving interior will tend to drag forward the slower-moving exterior,

and thus will increase its energy and make it recede from the sun. Thus the outer parts will be slowly expelled from the system. The inner parts, on the other hand, will have their motion delayed, and will therefore gradually fall into the sun. In time, therefore, the resisting medium will cease to exist. Diffusion would not affect the behaviour of a medium consisting of only one material; thermal conductivity would be continually transferring heat from the inside to the outside, but this would probably produce only a slight permanent change in the temperature distribution, and not a secular degeneration. If now ρ, μ, l, and τ be of the order of magnitude of the density, true viscosity, linear dimensions, and time of degeneration through viscosity, we have

$$\tau = \frac{\rho l^2}{\mu} \qquad \text{.........(1)},$$

from an analogy with the behaviour of other systems changing on account of viscosity, or from dimensional considerations since μ is of dimensions m/lt. Considering the motion of the medium within the orbit of Mercury, we can take

$$l = 6 \times 10^{12} \text{ cm.} \qquad \text{.........(2)},$$

$$\mu = 10^{-4} \text{ gm./cm. sec.} \qquad \text{.........(3)}$$

(since μ is independent of the density), and therefore

$$\tau \text{ is of order } 4 \times 10^{29} \rho \qquad \text{.........(4)}.$$

Consider next the motion of the planet Mercury, supposing it to have been moving in a highly eccentric orbit. The angular velocity about the sun of the matter near the planet's orbit being n, and the velocity of the planet relative to the medium being Υ, the resistance to the motion of the planet is $\frac{1}{2}\pi\rho a^2 \Upsilon^2$, where a is the radius of the planet*, since the relative velocity is much greater than the velocity of sound in the medium. Hence the time needed to reduce the relative motion to $1/e$ of its initial amplitude is of order $m/\frac{1}{2}\pi\rho a^2\Upsilon$, where m is the mass of the planet. Taking

$$m = 2 \times 10^{26} \text{ gm.}, \quad a = 2{\cdot}6 \times 10^{8} \text{ cm.}, \quad \Upsilon = 2 \times 10^{6} \text{ cm./sec.},$$

this time is

$$1000/\rho \qquad \text{.........(5)}.$$

If the whole extent of the medium was greater than, but of the same order of magnitude as, the distance of Mercury from the sun, the results of the last two paragraphs could be combined to give an estimate of the age of the solar system. For, if the time needed by the medium to disappear was short in comparison with the time required to produce a considerable effect on the eccentricity of the orbit of a planet, the medium would have gone before the eccentricity had been appreciably reduced, and the eccentricity would still be great. On the other hand, if the medium lasted much longer than the time required to affect the eccentricity considerably, the eccentricity would have been reduced practically to zero instead of only to about $\frac{1}{4}$. Hence these two times must be comparable.

* F. A. Lindemann and G. M. B. Dobson, *Proc. Roy. Soc.* **102 A**, 1922, 413.

Supposing them to be equal, we find that the density must have been of order 5×10^{-14} gm./cm.^{3}, and the time taken of order 2×10^{16} seconds or 6×10^8 years. On the other hand the actual medium must have had a much wider extent, and therefore must have been acting upon Mercury during the time required for a much larger medium to degenerate. If for instance matter from the distance of the orbit of Jupiter—to take what is perhaps an extreme hypothesis—passed across the orbit of Mercury and was ultimately absorbed into the sun, l in (1) would have to be taken equal to 8×10^{13} instead of 6×10^{12} cm., and thus τ would be of order $5 \times 10^{31}\rho$. Combining this with (5) we find that the time required would be of order 7×10^9 years, and the density would be 4×10^{-15} gm./cm.^{3}

The coefficient of true viscosity is independent of the density, and it has been shown that so long as the medium was affecting the planets, and we can suppose its motion dominated by the sun, its motion would remain the same. Hence however much the density declined, so long as the medium behaved like a gas, the rate of communication of angular momentum across any sphere within the fluid would remain the same; therefore the rate of absorption of matter into the sun or expulsion of matter from the system would remain the same. Hence in a further duration, comparable with that required for the medium to have its density reduced to half what it was initially, practically the whole of it would have disappeared. What remained would indeed have to be of density so low that the gas laws would not apply.

The only gaseous matter of sufficient density to be observable that exists outside the planetary atmospheres is that which reflects the zodiacal light, and it is natural to suppose that this is the last relic of the resisting medium. Its density, estimated from its luminosity*, is of order 10^{-18} gm./cm.^{3} Now the length of the mean free path of a hydrogen molecule at normal temperature and pressure is $1{\cdot}83 \times 10^{-5}$ cm., and in other circumstances is inversely proportional to the density. Hence in the zodiacal matter the mean free path is of order 10^9 cm., much less than the radial extent of this matter. Each molecule must therefore experience many collisions in every revolution around the sun, and this is the condition that the gas laws may apply. Thus the degeneration of the zodiacal matter must still be going on, and therefore it is probable that the whole age of the system is not more than twice the time needed to reduce the eccentricity of the orbit of Mercury to $1/e$ of what it was at the commencement. Hence the tidal theory of the origin of the solar system suggests that the age of the system is of order 10^9 to 10^{10} years. There are of course many sources of error in the data, the chief being in the primitive distribution of mass, but it is interesting to notice that the age obtained is of the same order of magnitude as the age of the earth inferred from the phenomena of radioactivity, which will be discussed later.

* Jeffreys, *M.N.R.A.S.* **80**, 1919, 139.

4·51. On the hypothesis that the density of the medium was 5×10^{-14} gm./cm.^{3}, the mass within the orbit of Mercury would be about 4×10^{25} grams, decidedly less than that of the planet. On the hypothesis that it was 4×10^{-15} gm./cm.^{3}, the mass within the orbit of Jupiter would be about 8×10^{27}, rather more than the mass of the earth, which is reasonable. If the density was 4×10^{-15} gm./cm.^{3} throughout the system, the mass within the orbit of Neptune would be comparable with that of Jupiter, which again is reasonable.

4·6. *Effects on other Planets.* It would be of interest to test the theory by application to the eccentricities of the orbits of the other planets, but unfortunately one meets with a serious mathematical difficulty. The gravitation of each planet would cause a condensation in the medium around it. The formula of Lindemann and Dobson, for the resistance of a gas to a body moving with a relative velocity much greater than that of sound, depends essentially on the condition that the matter around the body, right up to its surface, is continually changing. In the case of a gravitating planet this condition would hold only if the relative velocity of the gas towards it was greater than the velocity of escape from the surface of the planet; otherwise a portion of the gas would be retained by the gravitation of the planet. Thus a permanent gaseous condensation would be formed around the planet, and would be forced through the medium by it, so that the effective resistance to the motion of the planet would not be determined by the surface of the planet itself, but by that of this condensation. It seems reasonable to conjecture that the effective radius of such a condensation would be such that the velocity of a particle moving in a parabolic orbit about the planet at that distance (the disturbance due to the sun being ignored) is comparable with the velocity of the medium relative to the planet. If b denote this radius, m the mass of the planet, M that of the sun, and if the velocity of the planet relative to the medium is λ times the velocity rn of the medium itself, this gives

$$\left(\frac{2fm}{b}\right) = O(\lambda rn)^2 = O\left(\lambda^2 \frac{fM}{r}\right) \qquad \text{.........(1)},$$

whence

$$b = O\left(\frac{2mr}{\lambda^2 M}\right) \qquad \text{.........(2)}.$$

Taking as a preliminary standard the impossible case where the planet is at rest and therefore λ equal to 1, we find for the various planets the following values of b, in kilometres: Mercury 13; Venus 500; Earth 1000; Mars 130; Jupiter $1{\cdot}6 \times 10^6$; Saturn 8×10^5; Uranus $2{\cdot}5 \times 10^5$; Neptune $4{\cdot}4 \times 10^5$.

For the four terrestrial planets these numbers are much smaller than the actual radii, and the neglect of gravitation would therefore be justified if λ was equal to unity; thus in the early stages of their careers, when the eccentricities of their orbits were great, there would be little gravitational

condensation around them, and the effective surface would be practically the solid surface. But when the eccentricities became small, the value of λ would sink to something between the eccentricity and twice the eccentricity, and when this fact is allowed for, the only planet for which b is still less than the actual radius is Mercury. Hence the method of estimating the ages of the solar system already employed is unsuitable for any planet except Mercury. The values of b for the four great planets have always been greater than the radii.

4·61. For a non-gravitating planet the time required to produce a given change in the eccentricity is, by 4·5, proportional to $mr^{\frac{1}{2}}/\rho a^2 M$. For a great planet we may write the effective radius b for a. For Mercury $mr^{\frac{1}{2}}/a^2 M$ is $1 \cdot 2 \times 10^{-10}$ km.$^{-\frac{3}{2}}$; for Jupiter $mr^{\frac{1}{2}}/b^2 M$ is $1 \cdot 1 \times 10^{-11}$ km.$^{-\frac{3}{2}}$. Thus the effect of the gravitational condensation is so great that if the densities were the same and the eccentricity of the orbit of Jupiter large, the resistance would reduce the eccentricity of the orbit of Jupiter nine times as fast as that of Mercury. For smaller eccentricities a for Mercury would remain the same, while b for Jupiter would increase further. It would be dangerous to proceed far without more knowledge than we possess about the distribution of density in the resisting medium, but this effect of gravitational condensation offers at least a very striking suggestion as to the reason why the outer planets have small eccentricities, while Mercury has the largest in the system.

4·62. Another fact that is possibly related to the last is that the value of b for Jupiter, with λ equal to 1, is nearly the mean distance of the fourth satellite; for Saturn, about the distance of Titan; for Uranus, between those of Titania and Umbriel; and for Neptune, approximately the distance of its single satellite. Thus all the satellites except J VI, J VII, Iapetus, and the retrograde ones of Jupiter and Saturn would be within the gravitational condensations around their primaries almost from the start, and therefore their motions would have undergone the greatest disturbances from the resistance. The effects on the outer satellites would perhaps not become great until the reductions in the eccentricities of their primaries had considerably increased the sizes of the condensations. This may afford an alternative reason to capture for the wide gaps between the orbits of J IV and J VI, and between those of Hyperion and Iapetus.

4·7. So far no attempt has been made to account for the asteroids, which still offer an outstanding difficulty in this theory, as in every other. Their small size in comparison with Mercury indicates that, if they existed while the resisting medium was still exerting an appreciable influence on the terrestrial planets, their eccentricities must have been reduced to zero. Thus they must have been formed after the medium had almost disappeared, and therefore are not lost satellites. The fact that none of their

mean distances is much less than that of Mars or greater than that of Jupiter suggests that they are fundamentally related in origin; and the most natural explanation is that they were formed by the disruption of a primitive planet. Their total mass can hardly exceed a hundredth of that of the earth; this is much less than that of Mercury, but comparable with those of the great satellites of Jupiter, and the possibility that the asteroids were formed from a primitive planet which was itself a lost satellite of Jupiter may be entertained. The possible modes of rupture include explosion, rotational instability, and the tidal action of the sun or Jupiter. Rotational instability would only give a planet with one satellite comparable in size with itself; the same applies to the tidal action of the sun, for this would have to be magnified by resonance to lead to rupture, and then the theory that has been used to account for the origin of the moon would be applicable. Explosion is possible, though there is little evidence for it. If the planet contained enough radioactive matter to heat part of its interior up to the boiling point, the disruptive stresses might come to exceed the small gravitative power of such a mass and permit explosion, and if one explosion took place there is no reason why others should not follow, since gravity would diminish at every rupture. Tidal disruption by close approach to Jupiter is also possible; the small planet would have to approach very close to the surface of Jupiter, and might be repeatedly broken up during a single encounter. The relative velocities of the fragments would be comparable with that which would have enabled a particle at the surface of the small planet to escape from its influence, which is about 2 km./sec. The average departure of the orbital velocities of the asteroids from the mean of all is about 3 km./sec. The relative velocities might be in any direction, for the small planet might pass Jupiter considerably to the north or south of its orbital plane, and thus the orbits of the fragments might be considerably inclined, the possible inclinations again being of the same order as the actual ones. The aphelion distances of all would be almost equal to the distance of Jupiter when the encounter occurred.

The subsequent history of these bodies would be determined by such traces of the resisting medium as remained and by planetary perturbations. Their large eccentricities and inclinations agree equally well with the theories of explosion and tidal disruption. Some may have been captured; in particular it is possible that the satellites of Mars are captured asteroids. Large variations in the eccentricities of the orbits would be set up by the perturbations due to the planets, especially Jupiter, and the positions of the nodes and the apses would be continually varying. Thus the orbits would probably become considerably modified from their original form and position, and thus the fact that the smallest are contained wholly within the largest is consistent with the theory.

CHAPTER V

The Age of the Earth

"Is there any thing whereof it may be said, See, this is new? it hath been already of old time, which was before us" Eccles. i. 10.

5·1. Several methods of estimating the age of the solar system in general, and of the earth in particular, have been suggested. The very plausible hypothesis that the eccentricity of the orbit of Mercury has been reduced to its present moderate value by the action of a resisting medium has been utilized for this purpose in 4·5, and indicates that the age of the system is probably between 10^9 and 10^{10} years. The age thus found is the time since Mercury first took shape as a planet, probably a few years at most from the ejection from the sun of the matter that formed it, and perhaps before the disturbing star had made its closest approach to the sun. Several other methods have been suggested for determining various long intervals in the earth's history, but the intervals determined are not in all cases the same, and a little attention must be given to the probable extent of the differences between them arising simply from the fact that they are not all measured from the same event. It will be seen that what we do in estimating a long interval of time is to consider some change that has taken place according to a known law; if we know both the law and the extent of the change between two definite events, we can calculate the time that elapsed between them. In the present problem the later event is in every case the present time; the earlier depends on the process considered. We have seen that the earth, like all other planets, was probably initially fluid. It cooled to the solid state by radiation from the surface, and even after solidification a further time would elapse before the surface became cool enough for water to condense on it. At a still later epoch denudation and redeposition formed the first sedimentary rocks. At some stage during this process the moon was formed. An estimate of the time that has elapsed since any one of these events will give information about the time since any other, when we have some knowledge of the intervals between these early events in the earth's history.

5·2. The chief methods (in addition to the one based on the eccentricity of the orbit of Mercury) that have been suggested for the estimation of the various intervals called 'the age of the earth' are as follows:

1. The age of an igneous rock can often be found directly by means of the ratio of the quantities of Uranium and Lead in it, the rate of degeneration of Uranium to Lead following a known law. This is available for rocks of a very great geological age, but these rocks are intrusive into

still older sedimentary rocks, and therefore the ocean must be still older than the oldest rocks whose ages have been determined in this way.

2. The age of the ocean could be found directly if we knew the total amount of salt in it and the rate of transfer of salt to the sea by rivers.

3. We could similarly find the age of the ocean if we knew the total quantity of sedimentary rocks on the earth's surface and the rate of disintegration of igneous rocks.

4. If, as seems probable, the earth's surface has been maintained at nearly the same temperature throughout geological time, we can show that the sun must have been radiating energy at almost its present rate throughout that time. If we can find the total amount of energy the sun has radiated away, we can find an upper limit to the time it can have been radiating at its present rate, which gives an upper limit to the time needed to form all the rocks known to geologists.

5. The time since the solidification of the earth may be found if we know its law of cooling and certain facts about the initial and present distributions of temperature.

6. Tidal friction has probably increased the period of the earth's rotation, from the period of 4 hours mentioned in 3·2, to the present period of 24 hours. If we knew its rate we could find the time since the birth of the moon.

5·3. *Radioactivity.* By far the most satisfactory of these methods appears to be the first. Its history dates from the discovery by Becquerel, in 1896, that uranium salts gave out rays capable of producing an effect on a photographic plate enclosed in opaque paper. This effect was found to be independent of the physical and chemical states of the uranium present, and therefore it appeared to be a property of the uranium atom itself. Mme Curie carried out an elaborate investigation of the phenomenon, and found that the uranium ore used was much more active, in proportion to the amount of uranium present, than a pure uranium compound, and accordingly inferred that some other substance, still more active, was present. She succeeded in 1898 in isolating this substance, which proved to be a new element, and was given the name of Radium.

5·301. An astonishing fact was soon discovered about the occurrence of radium. It occurs in nature only in the presence of uranium, which itself never occurs without radium. The ratio of the masses of the two elements present in a sample of ore is almost always the same, except perhaps in some of the most recent rocks, namely $3{\cdot}4 \times 10^{-7}$ parts of radium to one part of uranium*. Such a constancy suggests a chemical combination, but the atomic weights of uranium and radium are respectively about 238 and 226, and therefore one atom of radium would have to unite with about three million atoms of uranium to give the

* Rutherford and Boltwood, *Amer. J. Sci.* 22, 1906, 1–3.

proper ratio. Such a complexity is not approached by the most complicated chemical compounds known, so that the chemical hypothesis is most unplausible.

5·302. A further discovery led to the explanation. Radium itself was found to undergo a gradual change. A mass of a radium compound enclosed in a sealed vessel was found to liberate a gas called 'radium emanation,' the rate of formation of the gas being simply proportional to the amount of radium present. The rate was such that, if initially one gram of radium was present, only half a gram would be present 1500 years afterwards. The rest would be transformed into the emanation and into the further disintegration products of the emanation. All uraniferous ores are many thousands of years old, on any geological hypothesis, and therefore we have to explain how it is that any radium exists at all: why it has not all broken up long ago. The explanation suggested by its invariable association with uranium is that as fast as it breaks up new radium is formed by the break-up of the uranium itself. The suggestion was experimentally verified by Soddy*, who prepared a specimen of uranium quite free from radium, kept it for some years, and was able to demonstrate the presence of radium in the specimen at the end of the experiment.

5·31. Uranium, however, does not pass straight to radium, nor is the emanation the final product. The latter, in fact, survives only a few days. Suppose then that u atoms of uranium are present at time t, and suppose that each atom of uranium becomes in succession unit amounts of various recognisable stages $X_1, X_2, X_3, \dots X_n$. Suppose the numbers of uranium atoms that have gone to form the amounts of these products present at the instant considered to be $x_1, x_2, x_3, \dots x_n$. Further suppose that what has been proved to be true of radium is true in general, namely that the rate of break-up of any product is simply proportional to the quantity present†, and accordingly that any product X_r generates in unit time $\kappa_r x_r$ units of the next product X_{r+1}. The rate of degeneration of atoms of uranium itself will similarly be denoted by κu. Then $u, x_1, x_2, \dots x_n$ satisfy the following differential equations:

$$\frac{du}{dt} = -\kappa u \qquad \text{.........(1)},$$

$$\frac{dx_1}{dt} = \kappa u - \kappa_1 x_1 \qquad \text{.........(2)},$$

$$\frac{dx_2}{dt} = \kappa_1 x_1 - \kappa_2 x_2 \qquad \text{.........(3)},$$

$$\dots\dots\dots\dots\dots\dots$$

$$\frac{dx_n}{dt} = \kappa_{n-1} x_{n-1} \qquad \text{.........(4)}.$$

* *Phil. Mag.* **9**, 1905, 768–79; **16**, 1908, 632–38; **18**, 1909, 846–65; **20**, 1910, 340–49.
† Rutherford and Soddy, *Phil. Mag.* **5**, 1903, 576–91.

Suppose that initially there are no degradation products present, so that when t is zero $u = u_0;\ x_1 = x_2 = x_3 = \ldots = x_n = 0$ ....(5).

The solutions of these equations are

$$u = u_0 e^{-\kappa t} \quad \text{.........(6)},$$

$$x_1 = \frac{\kappa u_0}{\kappa_1 - \kappa}(e^{-\kappa t} - e^{-\kappa_1 t}) \quad \text{.........(7)},$$

$$x_2 = \frac{\kappa\kappa_1 u_0}{\kappa_1 - \kappa}\left\{\frac{1}{\kappa_2 - \kappa}(e^{-\kappa t} - e^{-\kappa_2 t}) - \frac{1}{\kappa_2 - \kappa_1}(e^{-\kappa_1 t} - e^{-\kappa_2 t})\right\} \quad \text{......(8)},$$

and so on, the expressions becoming more and more complicated as later products are considered. If, however, all the degeneration products are short-lived in comparison with uranium, so that $\kappa_1, \kappa_2, \ldots \kappa_{n-1}$ are all great in comparison with κ, and t is so great that $1/t$ is less than the smallest of $\kappa_1, \kappa_2, \ldots \kappa_{n-1}$, the solutions reduce approximately to

$$u = u_0 e^{-\kappa t} \quad \text{.........(9)},$$

$$x_1 = \frac{\kappa u}{\kappa_1} \quad \text{.........(10)},$$

$$x_2 = \frac{\kappa u}{\kappa_2} \quad \text{.........(11)},$$

....

$$x_{n-1} = \frac{\kappa u}{\kappa_{n-1}} \quad \text{.........(12)},$$

$$x_n = \int_0^t \kappa_{n-1} x_{n-1}\, dt = u_0(1 - e^{-\kappa t}) \quad \text{.........(13)}.$$

Thus the amounts of all products present except the last remain in fixed ratios to each other and to the amount of uranium left, the ratios being such that the number of units of any product that break up in a given time is the same for all products and equal to the number of atoms of uranium that break up in that time. Thus we have an explanation of the constancy of the uranium/radium ratio. Further, the value of this ratio enables us to find κ. We know from experiments on the rate of disintegration of radium that every year 1/2280* of the radium present breaks up. If r is the number of radium atoms present in a rock specimen, and we allow for the difference in atomic weight between uranium and radium, we find (observing as is natural that each atom of uranium ultimately gives one atom of radium) that

$$\frac{r}{u} = \frac{3{\cdot}4 \times 10^{-7} \div 226}{1 \div 238} = 3{\cdot}6 \times 10^{-7} \quad \text{.........(14)},$$

whence $1/\kappa = 6{,}600{,}000{,}000$ years ....(15).

Knowing the rate of break-up of uranium, we shall now be able to find the time since the formation of any rock if we know the amounts of

* V. F. Hess and R. W. Lawson, *Wien. Sitzungsber.* 127, 1918, 1–55.

uranium and of the end product present. If indeed l denote the number of units of the end product present, we shall have

$$t = \frac{1}{\kappa} \log \frac{u + l}{u} \qquad \text{.........(16)},$$

and if l/u is small an approximation to this will be

$$t = \frac{l}{\kappa u} \qquad \text{.........(17)}.$$

Thus a chemical analysis of the rock should give its age when the end product is identified.

5·32. The argument so far given is independent of whether the various disintegration products are pure substances or not. All that has been assumed about their constitution is that each of them is made up of units of similar composition, each unit having been derived from one atom of uranium. The units themselves may be composed of atoms, which need not be all alike. Thus the occurrence of several chemically different substances in any disintegration product is possible. Experiments on the behaviour of uranium and radium have shown that this is actually the case. The gaseous emanation from radium contains two substances, namely the inert gas helium, which undergoes no further change, and a radioactive gas called niton; each atom of radium yields one atom of each of these gases. Niton again breaks up, each atom giving one atom of helium and one of a further transient substance called Radium A. The disintegration continues and no fewer than five atoms of helium are lost in succession from a single atom of radium. Now the atomic weight of radium is 226·4 according to the International Tables; Hönigschmid* finds 226·0. That of helium is 4. Thus the fifth product of the disintegration of an atom of radium should be five atoms of helium and one atom of some substance with an atomic weight of 206·4 or 206·0; or possibly the heavier product might be two similar or even dissimilar atoms. The direct identification of this substance, or these substances, by keeping a sample of radium until an analysable quantity of the end product has accumulated, has not yet been carried out, but indirect evidence has given very definite information about its nature. Before we proceed to this point, however, we notice that the atomic weight of uranium is 238·5 according to the International Tables, while Hönigschmid† finds 238·2. The difference between the atomic weights of uranium and radium is almost exactly three times the atomic weight of helium, and we therefore suspect that an atom of uranium loses three helium atoms before radium is formed. This is confirmed by the discovery of the two successive heavy metals, Uranium 2 and Ionium, the former produced by the loss of one helium atom from uranium, and the latter by the loss of another from Uranium 2.

* *Wien. Sitzungsber.* **120**, 1911, 1617–1652; **121**, 1912, 1973–1999, 2119–2125.

† *Wien. Anz.* **51**, 1914, 36–39.

The loss of a further helium atom by an atom of ionium produces radium. Thus the unit of the third disintegration product of uranium is an atom of radium and three atoms of helium, while the unit of the eighth product is eight atoms of helium and one or more atoms the sum of whose atomic weights is 206·0 to 206·5.

5·321. No element with an atomic weight differing from 206·0 to 206·5 by less than the probable error of an atomic weight determination was known when radioactivity was discovered. The nearest were Bismuth 208, Lead 207·1, and Thallium 204. If the product contained two similar atoms of atomic weight 103·2, the conditions would be satisfied. The elements whose atomic weights are nearest to this are Rhodium 102·9, and Ruthenium 101·7. The only one of these elements, and indeed the only element at all other than the known disintegration products, invariably found in uranium minerals is lead. We have therefore strong reason for believing that the final product of the disintegration of uranium is lead. The discrepancy in atomic weight still presented a difficulty, until direct determinations of the atomic weight of lead from uranium minerals were made in 1914 by Hönigschmid and Fräulein St Horovitz, Richards and Lembert, and Maurice Curie. It was found to be 206·2, not far from the predicted atomic weight, and almost a whole unit lower than that of ordinary lead. Thus the end product is identified; its unit consists of an atom of this new kind of lead, which will be called uranium lead, and eight atoms of helium. Like ordinary lead, uranium lead is not radioactive; no further degeneration occurs after this stage. Thus the determination of the age of a uranium mineral requires the determination of the amount of uranium still present, and of the amount of helium or of uranium lead present. When these are known, the ratio l/u is determinable, and then the age of the mineral can be found from 5·31 (17).

5·33. The use of the uranium/lead ratio for finding the ages of minerals was first attempted by Boltwood*, who found that the ratio of the amounts of uranium and lead present in uranium minerals of the same geological age was approximately constant. The uranium/helium ratio was applied in 1908–10 by the present Lord Rayleigh, then the Hon. R. J. Strutt. Both methods have been extensively used by Holmes. There is little doubt that the method involving the use of lead is the superior. It will be seen that the applicability of either method in any particular case depends on whether three conditions are satisfied. First, the final product estimated must have been absent from the mineral when this was first formed. There seems no reason to believe that original helium ever occurs in appreciable quantities in igneous rocks; original lead is common, but it is possible in many cases to attach a very high probability to its absence. Uranium in pitchblende is in the form of the oxide uranous uranate

* *American Journal of Science*, 23, 1907, 77–88.

U $(UO_4)_2$; no lead compound isomorphous with this occurs in these ores, and hence the lead and uranium must crystallize separately. When the crystals are too small for an analysis of a single crystal to be undertaken, as indeed is usually the case, it is more difficult to be sure that no crystals of a lead compound are intermingled with those of the uranium compound. If, however, we confine our attention to ores containing a large percentage of uranium, we can be practically certain that the amount of original lead is small compared with the amount of uranium. Doubt can in any case be dispelled or established by an atomic weight determination.

5·331. The second condition required is that radioactivity must be the only agency that has altered the composition of the mineral since it was formed. All the lead or helium generated in the mineral must still be in it. Thus minerals altered by heat or water must be excluded, since heat produces recrystallization and accordingly separation of lead from the associated uranium, and promotes the diffusion of helium through the rock or even into the free air, while water may produce a chemical separation. If the mineral becomes exposed to the air, loss of helium by diffusion into the air is certain, and even within the crust leakage into the surrounding rocks is probable. Thus a mineral that has not undergone metamorphosis by heat or water probably contains its proper amount of lead; but it is very doubtful whether any mineral contains the whole of the helium generated from its uranium. Thus estimates based on the helium/uranium ratio will be systematically lower than the true ages of the rocks.

5·332. Third, it must be possible to determine accurately the amount of lead or helium in the final product. The estimation of lead is not a difficult process, and presents no likelihood of serious error. In estimating helium, however, the mineral has to be ground to a fine powder, which then has to be heated *in vacuo* to drive off the included helium. Leakage occurs to some extent during the powdering process, and on this ground again the age found from the uranium/helium ratio must be too low.

We thus see that while, with proper caution in selecting the minerals to be examined, the uranium/lead ratio is likely to give correct determinations of the ages of rocks, the uranium/helium ratio is practically certain to give results systematically too low. Thus estimates made by the latter method can be regarded only as lower limits to the possible ages of the rocks they refer to.

5·34. The following table gives determinations by means of the uranium/lead ratio of the ages of minerals over a wide range of geological time. The data were utilized by Holmes to determine the ages in question, and it is mostly to him that the present status of the method is due. In his table* the value of $1/\kappa$ has been taken as $7{\cdot}5 \times 10^9$ years, whereas the

* *Discovery*, **1**, 1920, 118–23.

revised value of $6 \cdot 6 \times 10^9$ years obtained by Lawson and Hess has been adopted here. Holmes's numerical estimates have been reduced accordingly.

Table of Geological Periods and their Ages.

Era	Period	Age (Millions of years)	Era	Period	Age (Millions of years)
Quaternary	Recent	—	Primary	Permian	—
	Pleistocene	—		Carboniferous	260–300
Tertiary	Pliocene	—		Devonian	310–340
	Miocene	—		Silurian	—
	Oligocene	26		Ordovician	—
	Eocene	60		Cambrian	—
Secondary	Cretaceous	—	Archaean	Upper Pre-Cambrian	560
	Jurassic	—		Middle Pre-Cambrian	770–950
	Triassic	—		Lower Pre-Cambrian	1210–1340

5·35. Uranium and the elements derived from it are not the only radioactive substances known. The element thorium is radioactive, and produces helium in its degeneration just as uranium does, one atom of it liberating at least six atoms of helium in succession at a definitely ascertainable rate. By methods analogous to those used for uranium it has been found that one part of thorium in $1 \cdot 8 \times 10^{10}$ breaks up every year. It might therefore be thought that thorium, like uranium, could be used in the measurement of geological time. Unfortunately this is not the case. The atomic weight of thorium is 232·4 (International) or 232·2 (Hönig schmid), so that the loss of six helium atoms should leave something with atomic weight 208·4 or 208·2. This is almost the atomic weight of bismuth, but bismuth rarely occurs in thorium minerals. The only element whose atomic weight approaches this value that occurs regularly in thorium minerals is lead. Since uranium was found to yield an exceptional variety of lead, the possibility that thorium also gives an exceptional lead is to be considered seriously. A natural test to apply to the suggestion was to determine the atomic weight of lead in thorium minerals. This was done by Soddy and Hyman*, who found the mean value 208·4, in excellent agreement with prediction. Unfortunately, however, other investigations of the atomic weight of thorium lead have led to discordant results. The ratio of the amounts of uranium lead and thorium lead that would be expected to occur in a mineral can be calculated from the amounts of thorium and uranium present, the rates of decay of these two elements being known, and hence the theoretical atomic weight can be found. All other investigators have found that the atomic weight of lead in thorium minerals is less than is inferred from such a calculation†. The discrepancies are greater than the accuracy of the experiments would allow, and until they are explained the use of thorium minerals for measuring geological time cannot be considered reliable.

* *Trans. Chem. Soc.* **105**, 1914, 1402.

† For an account of the evidence relating to the end product of thorium, see Holmes and Lawson, *Phil. Mag.* **28**, 1914, 823–40, and **29**, 1915, 673–88.

5·36. Several explanations of these discordances have been offered, but no decision concerning their validity is yet possible. Holmes and Lawson are inclined to favour the hypothesis that the sixth degeneration product is actually a lead, but that it is itself radioactive and undergoes further slow degeneration. The difficulty in the way of such a hypothesis is that at least one further degeneration product should be formed, and should be still inside the mineral. Even if it were a gas, its high atomic weight would prevent it from being lost by diffusion as rapidly as helium, and it is known that uranium minerals habitually retain about half of their theoretical amount of helium. Thus the further product should be disclosed by a chemical analysis of the mineral, which is not the case. Presumably it should be a thallium or a member of the platinum group.

5·361. An alternative suggestion is that thorium is often associated with original lead. The association of uranium with original lead is unusual on account of the wide difference between the chemical properties of the two elements, but thorium and lead, being in the same column of the periodic classification, are more likely to crystallize together. Again, thorium minerals are usually found to have undergone metamorphosis since they were first formed. On any of these hypotheses one would be led to mistrust determinations of the ages of minerals by means of the thorium/lead ratio. It is indeed actually found that this ratio shows no recognizable relation either to the age found from the uranium/lead ratio in contemporary rocks or to the geological horizon ascertained by means of fossils.

5·37. The uranium/lead ratio forms the basis of a different method, due to Prof. H. N. Russell*. The proportion of uranium in the earth's crust is estimated as 7×10^{-6}, of thorium 2×10^{-5}, and of lead 22×10^{-6}. The fact that they are all of the same order of magnitude is of some incidental interest, seeing that lead is conventionally regarded as a common metal and the two putative parents as rare. The difference is one of accessibility rather than of quantity.

Now if all the lead were uranium lead, it would have resulted from 25×10^{-6} parts of uranium, and thus the original proportion of uranium would have been 32×10^{-6}. With our adopted rate of decay of uranium this makes the time required for the uranium to have been reduced to 7×10^{-6} equal to 11×10^{9} years. But the greater part of the lead in the crust is ordinary lead, and therefore has not all come from uranium. Thus this estimate of the original uranium, and therefore of the age of the crust, is too high.

Allowing for the decay of thorium, Russell finds that the lead of the crust could have been produced in 8×10^{9} years. The average atomic weight of such lead should be 206·9, a trifle below that of ordinary lead.

* *Proc. Roy. Soc.* **99**, 1921, 84–6.

This estimate is to be regarded as an upper limit to the age of the crust, since lead may have been present in the crust when it was first formed. Thus we may infer that the age of the crust is less than 8×10^9 years.

5·4. *The Denudational Methods*. We come now to the second and third methods of estimating geological time, which are usually described together as the geological methods; but since this may be held to constitute too narrow a use of the word 'geological,' they will be called the 'denudational' methods in the present work. Their methodological footing is altogether inferior to that of the method based on radioactivity. It has already been explained that an estimate of geological time requires two conditions: we must know the law satisfied by the rate of change we are using as our standard, and we must know its total extent in the interval we are measuring. The former condition is fulfilled by no denudational method; the latter is probably fulfilled with moderate accuracy by the accumulation of salt in the sea, but certainly not by the formation of sediments. Considering first the law of the change, we know the present rates of transport of salt and detritus to the sea by rivers, but it is not known how these rates have varied in the past. The rate of land erosion must depend on the slope of the land, the quantity, temperature, and carbon dioxide content of the rain falling, and on the nature of the soil exposed. No quantitative relation is known between the rate of denudation and any one of these factors, nor do we know even approximately how any one of these factors themselves has varied during geological time. We have some information relating to the type of rocks exposed in many places at various geological dates, but there is no place (except perhaps the bottom of the Pacific Ocean) where the nature of the solid surface has been the same at all geological dates, and there is no geological date such that the nature of the solid surface then is known for all points of the earth. We often know whether the land in some region was rising or sinking at a particular date, but we never know the precise extent of the elevation nor of the change of slope. Information concerning the amount and nature of the rainfall is still more vague in character. Finally, even if we had all this information, we should still not be able to find the rate of denudation at any geological date, since the physical laws connecting it with the relevant data are still unknown.

5·41. In the estimates hitherto made by the denudational methods*, it has always been assumed that the rate of denudation has been uniform throughout geological time, which is incorrect for the reasons just given. The amount of sodium carried to the sea annually is about $1{\cdot}56 \times 10^{14}$ grams, and the total mass of the sodium in the ocean is about $1{\cdot}26 \times 10^{22}$ grams. If the accumulation had been uniform, the age of the ocean would have

* Most of the following arguments are from *The Age of the Earth*, by A. Holmes, Harpers, 1913. Full references to earlier work on these lines will be found there.

been 8×10^7 years. This is practically Joly's estimate. But much of the sodium carried to the sea is derived from the denudation of sedimentary rocks, and has therefore been in the sea before. Igneous rocks contain only about 2 per cent. of the chlorine required to combine with their sodium, and therefore it is probable that nearly all the chlorine in the ocean is of volcanic origin. If so, the amount of sodium corresponding to the amount of chlorine carried to the sea by rivers must be almost wholly derived from sedimentary rocks. This amounts to about 60 per cent. of the whole amount of sodium carried by rivers. Hence the amount of new sodium is unlikely to exceed $6 \cdot 9 \times 10^{13}$ grams annually. If this value is adopted, the corresponding age of the ocean is $1 \cdot 8 \times 10^8$ years. This is still too low, for much unchlorinated sodium must also be included in the sedimentary rocks, so that some of even the unchlorinated sodium must have been in the ocean before. Hence a further increase by a practically incalculable amount is necessary.

5·42. The method based on the accumulation of sediments also meets with additional difficulties. After an elaborate discussion of the possible ways of utilizing it, Holmes decides that the most satisfactory is probably as follows. The igneous rocks at present exposed at the earth's surface produce altogether a cubic mile of sediments in five years. The total volume of sediments on the earth's surface is estimated at 7×10^7 cubic miles. Sediments derived from other sediments are not new, and are excluded by this method. The age of the ocean is thus estimated at $3 \cdot 5 \times 10^8$ years.

5·43. These two methods are, however, incapable of giving satisfactory determinations of the age of the ocean, for the reasons already given. The results are much smaller than the age of the oldest known rocks whose uranium/lead ratios have been determined, the difference being too great to be explicable by uncertainty as to the present rate of denudation or the actual total extent of denudation. Accordingly they are to be regarded as in error owing to variations in the rate of denudation. Their value, such as it is, is that they amount to a proof that the present rate of denudation is several times greater than the average of the past; they are not estimates of the age of the ocean.

5·5. *The Solar Energy Method.* The fourth method is the original method of Lord Kelvin. If m be the mass of a body and U the gravitational potential at its surface, the kinetic energy acquired by a mass dm in falling from an infinite distance to the surface of the body is Udm. When the added mass reaches the surface the kinetic energy becomes converted into heat and hence becomes available for radiation. Thus the total energy liberated by condensation can be found by supposing the whole mass to be brought up gradually from an infinite distance and deposited in thin uniform layers over the surface, and adding up the energies acquired by all in their fall. If the

mass be supposed of uniform density ρ in its final stage, and if the radius at any intermediate stage be r, and the final radius a, we find for the energy of condensation the amount

$$\begin{aligned} W &= \int \frac{fm}{r}\,dm = \int_0^a \tfrac{4}{3}\pi f\rho r^2 \,.\, 4\pi\rho r^2\,dr \\ &= \tfrac{16}{15}\pi^2 f\rho^2 a^5 \\ &= \frac{3}{5}\frac{fM^2}{a}, \end{aligned}$$

where M is the final mass. In the case of the sun this amounts to $2{\cdot}6 \times 10^{48}$ ergs, or $1{\cdot}3 \times 10^{15}$ ergs for each gram of the sun's mass. The latter result shows that any chemical energy in the sun must be of very small importance in comparison with condensational energy, since the energy of the most violent chemical reactions amounts only to quantities of the order of 10^4 calories per gram, or of 10^{11} ergs per gram.

Now the radiation received by a square centimetre of material exposed normally to the sun's radiation at the earth's distance from the sun is 0·03 cal. per second. Taking the earth's distance from the sun as $1{\cdot}5 \times 10^{13}$ cm., we see that the sun must be losing energy at a rate of $3{\cdot}3 \times 10^{33}$ ergs per second. Thus the total condensational energy of the sun would provide for radiation at the present rate for $7{\cdot}8 \times 10^{14}$ seconds or $2{\cdot}5 \times 10^7$ years.

5·6. *Possible Source of Solar Energy.* This estimate is only of the order of a fiftieth of the age of the oldest rocks, as found from the uranium/lead ratio. Accordingly either some other, and much more abundant, supply of energy is available in the sun, or else there is a definite inconsistency between two results both based on physical laws. Numerous attempts have been made to discover such a source of energy, but hitherto none has been proved adequate. The most probable appears to be one of Jeans, developed by Eddington. On the theory of relativity all forms of energy possess mass. The mass of a quantity of energy W is indeed W/c^2, where c is the velocity of light. Thus the continual loss of energy from the sun must have lowered its mass. Jeans, however, went further by suggesting that much of this energy was derived, not from the sources of energy already known, but from the disappearance of part of what had hitherto been considered the invariable intrinsic mass. Strictly speaking, such a suggestion amounts to nothing more than a statement in another form of what we knew already, namely that if the sun has actually lost so much energy, it must have lost a corresponding mass. Its novelty consists in the hint that the loss of mass, and not the loss of energy, is the more fundamental. About the same time Aston was investigating the properties of isotopes. The discovery that uranium lead, thorium lead and ordinary lead differed from one another in their atomic weights, but were quite indistinguishable in their chemical properties, was the first instance known of such a property. Aston's work, however, has shown that a large fraction, perhaps the majority, of the

known elements are not simple substances, but mixtures of two or more substances capable of being distinguished by their atomic weights, but not by chemical means. It is possible that ordinary lead is itself a mixture of uranium lead and thorium lead, but its analysis has not yet been achieved. Now Aston found that the atomic weights of all elements that do not break up in his mass-spectrograph are whole numbers; and that wherever the international atomic weight, oxygen being taken as 16, is not a whole number, the element is always a mixture of two or more isotopes whose atomic weights are whole numbers. The only exception is hydrogen, whose atomic weight is 1·008, and which is a simple substance.

5·61. Now in the development of stars it is possible that transmutation of elements takes place on an enormous scale. If the loss or gain of energy in any such transformation is insufficient to change the sum of the atomic weights of the participating atoms by a perceptible fraction of the unit, and if all elements have been formed by the union of atoms of some one simpler element, it would be expected that their atomic weights would differ among one another by multiples of the atomic weight of this primary element. This is not the case; helium is the lightest element whose atomic weight is a whole number (namely 4), and when the elements are arranged according to their atomic weights, the atomic weights of consecutive elements usually differ, not by 4, but by 1, 2, or 3, indicating that they are not formed by simple addition of helium. The alternative is that they are really composed ultimately of hydrogen, but that each hydrogen nucleus, when it combines with another hydrogen nucleus or with another atom to form a heavier atom, loses enough mass to make it increase the atomic weight of the other atom, not by its own full weight 1·008, but by exactly 1. If then a star was originally composed of hydrogen, and this aggregated afterwards to form heavier elements, 0·008 of its mass would have been lost in the process. In particular, the possible loss of mass of the sun would be $1{\cdot}6 \times 10^{31}$ grams. The corresponding loss of energy, which evidently must be the energy liberated by the union of the hydrogen atoms, amounts to $1{\cdot}6 \times 10^{31}c^2$ ergs, or $1{\cdot}4 \times 10^{52}$ ergs. This would supply energy at the present rate for $1{\cdot}4 \times 10^{11}$ years, a longer interval than has yet been indicated by any estimate of geological time. Thus the problem of the supply of energy is apparently solved; but some difficulties remain.

5·62. The sun is a dwarf star, that is to say, a star of high density and probably largely liquid or solid. The usual evolution of a star consists of a condensation from a highly distended and cold state, up to a very hot stage, with a density of the order of $\frac{1}{4}$, at which liquefaction begins, followed by a further condensation with fall of temperature. It is at the hottest stage that hydrogen and helium are predominant. If then the aggregation of hydrogen atoms is the chief source of the energy of a star after it has passed its maximum temperature, this energy must apparently have been

stored up during the giant stage, and we again require to know where it came from. Condensation, by what has already been shown, is quite inadequate. It remains possible that only a small fraction of the hydrogen in the star actually undergoes aggregation, in which case the hypothesis will not require the disruption of heavier atoms to give hydrogen during the giant stage.

5·63. A relevant datum is supplied by the variable star δ Cephei. It has been seen that, given the mass and radius of a star, we can find its condensational energy (apart from a correction due to heterogeneity, which will not affect the order of magnitude). If in addition we know the rate of emission of energy (in other words, the bolometric absolute luminosity) we can find how fast the star should be contracting if the whole output of energy is derived from condensation. Now the star in question executes a regular pulsation, whose period must theoretically be proportional to the inverse square root of the density. Thus condensation should alter the period at a calculable rate. This theory of Cepheid variation is due to Prof. Eddington* and Prof. Shapley†. The former estimates that the period of δ Cephei should decrease by about 40 seconds annually if its energy came wholly from condensation. The actual change, if any, was estimated by Chandler as 0·05 second annually. It follows that some other source of energy must be available.

5·64. Eddington's comparison, however, refers only to giant stars. So far no observational information from astronomy is available to permit a corresponding comparison for dwarf stars. Eddington's result amounts to strong evidence that some source of energy much more powerful than condensation is accessible to giant stars, if we argue to the dwarf stars by analogy there will be no difficulty in accepting the time estimates given by the uranium/lead ratio. But the coordination of the varied data remains very imperfect.

The two remaining methods will be described in later chapters, the one gives only a lower limit, the other the order of magnitude, of the intervals they measure.

5·7. Thus of all the methods suggested for the measurement of the age of the earth, only the uranium/lead ratio method is quantitatively satisfactory. It gives only the age of the oldest known igneous rocks, which are intrusive into conglomerates that must have been formed from still older rocks by sedimentation. Hence we can assert that the age of the ocean exceeds 1340 million years. It is now necessary to discuss the lengths of the various stages in the history of the earth that elapsed before the formation of the ocean.

* *M.N.R.A.S.* **79**, 1918, 2–22.

† *Ap. J.* **40**, 1914, and **48**, 1918, *Mount Wilson Contributions*, No. 92 and Nos. 153–4.

5·8. *The Early History of the Earth.* It is known, from a comparison with Capella, that when the sun was a giant star it must have radiated at about 13 times its present rate. Thus the earth was receiving thirteen times as much radiation as it does now, and its surface temperature must have been correspondingly higher. The condition that the surface of the earth should have remained at a steady temperature is that the heat lost by radiation should have been just enough to balance that received from the sun. The rate of radiation from the earth being proportional to the fourth power of the absolute temperature, it follows that while the sun was a giant star the temperature of the earth's surface, if steady, must have been $(13)^{\frac{1}{4}}$ times what it is now. The mean absolute temperature of the earth's surface is now about 280°. Hence when the sun was a giant star the surface of the earth must have been at 530° absolute, or 257° Centigrade. If the temperature was not steady, it would be owing to cooling of the earth, and then the surface would be still hotter owing to the supply of heat transferred from the interior. Hence no water could condense on the earth until the sun had passed the giant stage.

Before an ocean could form, the surface temperature of the earth must have sunk below the boiling point of water, 373° absolute. Solar radiation would maintain this if it had three times its present intensity. Now if the radiating surface of the sun was no larger than at present, such radiation would be maintained if the sun's effective temperature was 8200°, corresponding to spectral type F0. If the sun was more distended than at present, as it certainly would be, a still lower temperature would be necessary to keep terrestrial water from condensing. Combining this result with that of the last paragraph, we see that the ocean could not have been formed until the sun had gone through the giant stage and was past the early stage on the decline. Geological time therefore commenced when the sun was already a yellow dwarf star.

5·81. The sun is now a dwarf of type G0. Thus geological time corresponds to only a part of the time the sun has taken in passing from the dwarf F0 stage to the dwarf G0 stage, which one might expect to be only a small fraction of the whole time taken by the sun in passing from the giant M stage to its present state. This is not necessarily the case. Eddington has shown* that the observed rate of change of the period of δ Cephei corresponds to a change from the giant G to the giant F stage in about 10^7 years. This star is much more massive than the sun, and accordingly it appears probable that the sun did not take longer than this to pass through the whole of the giant stage. Information about the rate of evolution in the dwarf stage is more indefinite; but it is significant that of the 105 stars within 10 parsecs of the sun, only 17 are of earlier spectral type than the sun†, suggesting that the further advanced a dwarf star is

* *M.N.R.A.S.* **79**, 1918, 19. † W. J. Luyten, *Harvard Annals*, **85**, No. 5.

in its evolution the more slowly its type changes. Some of these stars are probably in early stages, on account of their great mass, and some of those in later stages are probably in those stages because their small mass is inconsistent with a temperature as high as the present temperature of the sun. It is, however, probable on the evidence that the time spent by a star in the giant stage, and in the dwarf stage before reaching type F0, is only a small fraction, perhaps about 1/10, of the time taken to pass from type F0 to type G0. If this hypothesis, which is supported by what little evidence we possess, is correct, the ocean must have existed for much the larger part of the history of the earth.

5·82. The time taken by the earth in liquefying and solidifying is more easily shown to be a small fraction of its age. The rate of loss of heat per gram of its mass is given by the formula of 2·5, namely $3\sigma V^4/\rho a$, where a is the radius. If we take

$$\sigma = 5 \times 10^{-5} \text{ c.g.s. Centigrade units,}$$
$$V = 3000°,$$
$$a = 6{\cdot}4 \times 10^8 \text{ cm.,}$$

so that we are taking practically the boiling point of terrestrial materials and the present radius, this amounts to 4 ergs per second per gram, or 3 calories per gram per year. The total energy due to the initial temperature of the gaseous earth could hardly exceed 6000 calories per gram, while the formula in 5·5 for the energy due to condensation shows that the energy per unit mass is proportional to M/a, and therefore in the case of the earth must have been about 10,000 calories per gram. Thus if the earth remained gaseous the whole of the condensational and initial thermal energy would have been radiated away in 5000 years at most. The actual rate of radiation would be somewhat greater than this, since the earth in the gaseous state must have been more distended and therefore have had a larger radiating surface than at present. Thus liquefaction must have been complete within 5000 years of the formation of the earth.

If we suppose the primitive earth to have had within itself a store of energy per unit mass comparable with that suggested for the sun in 5·61, this energy would maintain radiation at the rate appropriate to a gaseous earth for 1/33 of the time it enabled the sun to radiate at the same temperature; but the present rate of production of heat from atomic changes in the earth suggests that the atomic energy per unit mass in the earth is in no way comparable with what appears to be required for the sun.

5·83. The liquefaction would be complete when all the latent heat of evaporation had been radiated away, and cooling to the melting point would then proceed. The latent heat of fusion would then have to be lost before solidification was complete. The total loss of heat from the commencement of condensation to the completion of solidification would hardly exceed 2000 calories per gram. Taking the melting point as 1500°,

we find that the rate of loss of heat per gram at that temperature would be $\frac{1}{5}$ calorie per year. Thus the time between the onset of liquefaction and the advancement of solidification to such a stage that internal convection was stopped could not exceed 10,000 years. Hence we can allow in all 15,000 years from the formation of the earth to its solidification to this extent.

The formation of the moon can be placed with some certainty with regard to the process just outlined. It has been seen that conditions suitable for the formation of the moon by resonance did not exist till the diameter of the earth had almost attained its present value: until that time rotation would be too slow for resonance. Hence it could not have taken place until liquefaction was nearly complete, at the earliest. The conditions could not occur after the solidification had proceeded far enough to stop internal fluid motions, for such solidification would introduce a good deal of rigidity (see p. 112), which is absent from a fluid, and thereby would considerably shorten the period of a free vibration. Thus the free vibration would again become quicker than the period of the semidiurnal tide; in other words, rotation would again be too slow for resonance. Hence we can say that the birth of the moon took place while the earth was almost wholly liquid, the amounts of gaseous and solid constituents present being insufficient to produce any considerable effect on its period of free vibration. The moon existed before the ocean, and was probably formed within 10,000 years of the formation of the earth as a separate body.

5·9. *Summary*. The earth probably became solid within 15,000 years from its ejection from the sun. So long as the sun remained a giant star, the surface of the earth could not become cool enough for an ocean to condense; thus a considerable time, perhaps comparable with the whole time from the original rupture to the present, may have elapsed between the solidification of the earth and the condensation of the ocean. From astrophysical considerations, however, it seems more probable that the time between these events was a small fraction of the whole age of the earth. The moon, if it was ever part of the earth, was formed about the time when solidification was starting. The interval from the formation of the ocean to the present time constitutes the greater part of the age of the earth, and it is possible that all the previous stages together occupied only an insignificant fraction of it.

Collecting our results, we have found the following:

1. From the eccentricity of the orbit of Mercury, we saw that the whole time since the rupture is probably between 10^9 and 10^{10} years.

2. The ages of the oldest known minerals, found from the uranium/lead ratio, are about $1{\cdot}3 \times 10^9$ years, and since the geological evidence indicates that some sedimentary rocks are still older, the age of the ocean must exceed $1{\cdot}3 \times 10^9$ years.

3. The amounts of uranium, thorium, and lead in the crust as a whole indicate a time less than 8×10^9 years since solidification

The second and third results are in good agreement with the first. The age of the earth is therefore probably between $1{\cdot}3 \times 10^9$ and 8×10^9 years. The other methods are less satisfactory. The best inference that can be drawn from the denudational methods is that the present rate of denudation is about four times the average of the past, a conclusion in harmony with the fact that the present time is just after a glacial period and a period of mountain formation, both of which would tend to increase the rate of denudation very considerably. The Kelvin method, based on the sun's supply of energy, leads to inferences about the stars that do not accord with the facts, and it therefore cannot be trusted. It indicates rather that the sun has some source of energy other than condensation; and it is possible that the union of hydrogen atoms to form heavier elements supplies most of this energy.

CHAPTER VI

The Thermal History of the Earth

"I know it's something humorous, but lingering."

W. S. Gilbert, *The Mikado*.

6·1. *Mode of Solidification.* The solidification of the earth probably took place in the manner indicated by Lord Kelvin*. Most known rocks contract in freezing. Thus the first step in the solidification of a liquid globe would be the formation of a thin superficial crust of higher density than the inside; this would be unstable, and would therefore break up and sink until melted again. A further crust would then form, and the process would be repeated. Ultimately so much of the heat of the interior would have been used up in fusing the sinking solid fragments that it could no longer melt them. They would therefore remain solid, and would accumulate until at last the mass consisted of a honeycombed solid, and convection ceased.

6·11. Now as the newly formed solid sank, it would be exposed to greater pressure, and its temperature would rise through compression. If it rose by more than the melting point would be raised by the same increase of pressure, the rock would melt again and cease to sink. Thus the rate of increase of temperature downwards in the primitive crust could not exceed the rate of increase on the hypothesis that the rocks at every depth were at the melting point appropriate to the pressure there. The change of melting point due to an increase of pressure is given† by

$$\frac{\partial T}{\partial p} = \frac{(v_2 - v_1)\,T}{L} \qquad \text{.........(1)},$$

where T denotes the melting point on the absolute scale of temperature, p the pressure, $v_2 - v_1$ the change in volume of a gram of the substance on melting, and L the latent heat of fusion. The melting points of rock materials are usually over 1000° C., or say 1300° absolute. Accurate information about the change of volume of a rock material on melting seems to be lacking, the best being that of Day, Sosman, and Hostetter‡, who estimate that diabase in changing from the crystalline state to the glassy state expands about 10 per cent. of its volume. The further expansion from the glassy to the liquid state is probably less. If the density is 3, this makes $v_2 - v_1$ equal to 0·033 c.c. per gram.

If we take the latent heat of a mineral melting at 1000° C. to be 100 cals./gm., the above formula gives $\frac{\partial T}{\partial p} = 8° \times 10^{-9}$ per dyne/cm.^{2} The

* Kelvin and Tait, *Treatise on Natural Philosophy*, Part 2, 482.

† Cf. O. Sackur, *Thermochemistry and Thermodynamics*, Macmillan, 1917, 221.

‡ *Amer. Journ. of Science*, 37, 1914, 1–39.

increase of pressure due to a descent of 1 cm. in rock of density 3 gm./cm.^{3} is 3000 dynes/cm.^{2}, and therefore if x denotes the depth we have

$$\frac{dT}{dx} = 2^\circ{\cdot}5 \times 10^{-5}/\text{cm.} \qquad \text{.........(2).}$$

This is an upper limit to the initial rate of increase of temperature inwards, for, as has just been explained, if the initial rate exceeded this, the rock would melt in its descent and cease to sink.

6·12. The centre of the earth is probably composed of metallic constituents much denser than the rocks of the crust*, and therefore the fragments of the crust would not sink into the core. Equality of temperature would be established across the boundary of the two types of material, but there would be no convective interchange between them.

Accordingly we may suppose that the initial temperature of the earth was of the form $S + mx$, where S is the melting point at zero pressure of the typical rocks of the crust, x is the depth, and m is less than $2{\cdot}5 \times 10^{-5}$ degrees per cm. It will be supposed constant in the following work.

6·13. The primitive earth would be a honeycombed solid, the cells being still filled with liquid. Heat would be gradually conducted out of these, and solidification would thus proceed till complete. Meanwhile the surface would be losing heat by radiation and absorbing radiation from the sun. When the sun had cooled sufficiently to permit an ocean to form the surface temperature would probably be determined almost wholly by the balance between solar radiation absorbed and terrestrial radiation emitted; the supply of heat to the surface by conduction from the interior would be too slow to affect the surface temperature appreciably. The evidence is against any systematic change of solar radiation in one direction during geological time, and in the present theory of the cooling of the earth it will be assumed that the surface temperature has remained constant ever since solidification. This postulate does not hold for the interval before the formation of the ocean, but reasons have been given for believing that this is a small fraction of the whole history of the earth, and the error thus introduced will not be considerable. It will also be assumed that the whole of the crust has been solid throughout the time considered; in other words, any isolated fluid regions enclosed among the solid will be ignored.

6·2. *The Cooling of the Crust.* The problem of the flow of heat in a uniform isotropic solid depends on the differential equation

$$\frac{\partial V}{\partial t} - \frac{k}{c\rho}\left(\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}\right) = \frac{P}{c\rho} \qquad \text{.........(1),}$$

* See later, Chapter XIII.

where V is the temperature, t the time, x, y, and z the three Cartesian position coordinates, k the thermal conductivity, ρ the density, c the specific heat, and P the rate of generation of heat per unit volume by radioactivity or chemical change. In the present problem the cooling is supposed to start at the surface and to spread gradually downwards, and it will be supposed that it has had time to affect only a layer whose thickness is small in comparison with the radius of the earth. Accordingly if the axis of x be vertically downwards $\frac{\partial^2 V}{\partial y^2}$ and $\frac{\partial^2 V}{\partial z^2}$ will be small in comparison with $\frac{\partial^2 V}{\partial x^2}$, and the differential equation reduces to

$$\frac{\partial V}{\partial t} - h^2 \frac{\partial^2 V}{\partial x^2} = \frac{P}{c\rho} \quad \text{.........(2)},$$

where h^2 has now been written for $k/c\rho$.

Let some particular integral of this equation be V_0. As a rule P is, with sufficient accuracy, a function of x alone, so that a particular integral is $-\iint P/k\,dx\,dx$. Whether this is so or not, however, provided only that a particular integral can be found, we have still to find a function V_1 such that when t is zero $V_0 + V_1$ is equal to the specified initial value of the temperature at any depth, such that when x is zero $V_0 + V_1$ is equal to the actual surface temperature at any time different from zero, and such that V_1 satisfies the differential equation

$$\frac{\partial V_1}{\partial t} - h^2 \frac{\partial^2 V_1}{\partial x^2} = 0 \quad \text{.........(3)}.$$

If such a function is found, $V_0 + V_1$ will be the solution of (2) required.

Subject to the condition that cooling has not had time to become appreciable at depths comparable with the radius of the earth, the solution will be almost the same as if the solid was infinite in depth. The constant surface temperature will be taken for the zero of temperature.

If V_0 and the initial distribution of temperature are known, the value of V_1 for all values of x when t is zero can be found. Let this be $f(x)$. Then it can be shown* that the solution is

$$\begin{aligned} V_1 &= \frac{1}{\sqrt{\pi}} \int_{-\lambda}^{\infty} e^{-q^2} f(2qht^{\frac{1}{2}} + x)\, dq - \frac{1}{\sqrt{\pi}} \int_{-\infty}^{-\lambda} e^{-q^2} f(-2qht^{\frac{1}{2}} - x)\, dq \\ &= \frac{1}{\sqrt{\pi}} \int_0^{\infty} \{e^{-(q-\lambda)^2} - e^{-(q+\lambda)^2}\} f(2qht^{\frac{1}{2}})\, dq \quad \text{.........(4)}, \end{aligned}$$

where λ has been written for $x/2ht^{\frac{1}{2}}$. This is the solution of the most general problem of the class considered. Since $f(x)$ is now a known function, it follows that V can be found for any values of the depth and the time. It will be seen, however, that in addition to the thermal constants of the rocks, some knowledge of the form of the function P is required before the method can be used for numerical evaluation of the temperature. We

* Ingersoll and Zobel, *Mathematical Theory of Heat Conduction*, Ginn, 1913, 76; Riemann, *Partielle Differentialgleichungen*, 1869, 123–4.

have no direct quantitative knowledge of the radioactivity of rocks, except close to the surface; it will be found, however, that the knowledge available is such that a wide range of variation in the form of P makes little difference in the distribution of temperature inferred.

6·21. Let us consider first the value of P in the neighbourhood of the surface. This can be determined from analyses of actual rock specimens. It has been explained already that radioactive elements in their degeneration liberate helium. This comes off in the form, not of neutral helium atoms, but of α particles, which are helium atoms without their two outer electrons. The velocities of emission of these particles have been measured. It is found that all α particles emitted by the same element have the same velocity. The number emitted per second per gram of radium can be found by direct counting, and the mass of an α particle is known. Hence the total mass and the velocity of the particles emitted are known, and therefore their kinetic energy can be found. Again, it has already been explained that in any actual rock the quantities of all the radioactive substances present are such that the same number of atoms of each break up every second. Hence if we know the amount of any one radioactive element in a rock we can find the rate of production of α particles by each member of the series. The velocities of the α particles from all members being known, the total kinetic energy of those from each element can be found, and hence finally if the quantity of any one radioactive element of the Uranium series in a rock is known, the kinetic energy of all the α particles emitted per second by all the members of the Uranium series present can be calculated.

6·211. Now in an actual rock an emitted α particle cannot proceed far before it is stopped by the surrounding material. Its kinetic energy then becomes converted into heat. Thus the presence of a known quantity of radioactive materials in a rock enables us to infer the rate of supply of heat to that rock. Indeed, one of the first facts noticed about radium was that its temperature was always a trifle above that of its surroundings. The result is that for every gram of uranium present, uranium and its products produce $8{\cdot}0 \times 10^{-5}$ calories per hour. This result requires to be increased somewhat to allow for the fact that α particles are not the only form of radiation from radioactive substances. In addition they send out β particles, which are free electrons, and γ rays, which are electromagnetic waves closely resembling X-rays. The energy liberated by the absorption of these in the medium is enough to increase the estimate just mentioned to $8{\cdot}5 \times 10^{-5}$ calories per gram per hour. This has been confirmed by H. H. Poole*, who filled a Dewar flask with pitchblende, kept the outside at a fixed temperature, and determined the difference in temperature between the inside and the outside by means of a thermocouple. He

* *Phil. Mag.* **19**, 1910, 314–326.

found the total rate of emission to be 10^{-4} cal./gm. hr. The former estimate is, however, probably the more reliable, since it is a task of extreme difficulty to measure the small temperature difference ($0^\circ\cdot007$) involved to the degree of accuracy required. The rate of evolution of heat from thorium and its products is $2\cdot2 \times 10^{-5}$ cals./gm. hr.

6·212. The amounts of radium and thorium in rocks are found by separating their emanations and measuring the ionisation produced by the α radiations from these. Thus the number of helium atoms produced per second from any rock is measured directly, and the rate of evolution of heat is therefore calculable. Many determinations of the supply of heat to rocks through radioactivity have been made in this manner, and the results have been systematized by Holmes*. The following table, for plutonic rocks, is due to him.

	Acid Rocks	Basic Rocks	Average
Heat produced from Uranium series (cals./cm.^{3} sec.)	$4\cdot3\times10^{-13}$	$1\cdot6\times10^{-13}$	$3\cdot0\times10^{-13}$
„ „ „ Thorium „ „ „	$5\cdot8\times10^{-13}$	$1\cdot1\times10^{-13}$	$3\cdot4\times10^{-13}$
Total	$10\cdot1\times10^{-13}$	$2\cdot7\times10^{-13}$	$6\cdot4\times10^{-13}$

It is seen that although one gram of thorium produces less heat in a given time than one gram of uranium, this is more than compensated by the larger quantity of thorium present in the earth's crust.

6·213. On inspecting these results, we notice that the radioactivity of basic rocks is systematically lower than that of acid rocks. Now basic rocks are on the whole denser than acid ones, and therefore may be expected to occur at lower levels in the crust. Further, the more basic the rock considered is, the denser and the less radioactive it is found to be. We may therefore suspect that the radioactivity of rocks diminishes as the depth increases. This is strikingly confirmed by a discovery of Lord Rayleigh†. The rate of leakage of heat from the earth is the product of the thermal conductivity into the rate of increase of temperature downwards at the surface. The thermal conductivity at the surface is about 0·006 c.g.s., and the vertical temperature gradient, observed in borings, is $0^\circ\cdot00032$ C. per cm. Thus the rate of leakage of heat from the earth is $1\cdot9 \times 10^{-6}$ cals./cm.^{2} A depth of 19 km. of standard acid rock would therefore be capable of supplying all the heat reaching the surface of the earth from within. If the rocks below a depth of 19 km. were as radioactive as this, more heat would be produced than is lost, and the interior of the earth would be getting hotter. This result is not acceptable in view of our belief that the earth is cooling; but the argument is capable of a more decisive formulation.

6·214. Supposing that the radioactivity falls off with increasing depth, we may suppose that the actual law will lie between two extremes. On the first alternative, radioactivity will be uniform down to a certain depth, at present unknown, and zero at all greater depths. On the second alterna-

* *Geological Magazine*, Feb. 1915, 60–71. † *Proc. Roy. Soc.* 77 A, 1906, 475.

6–2

tive, it will decrease exponentially, starting from the surface. It will be seen that in either case the distribution of temperature with depth at any time can be determined, subject to a knowledge of the relevant numerical constants. It will be found that the only constant not known is, in the first case, the depth of the radioactive layer, and in the second, the constant in the exponential involved. In either case the known temperature lapse rate at the surface provides one quantitative datum, which is sufficient for the determination of this remaining unknown.

6·3. *Uniformly Radioactive Layer of Finite Depth.* Taking first the uniform distribution in a surface layer, let us suppose that the depth of the layer is H; that P is constant in this layer and equal to A, and that at depths greater than H, P is zero. Then a particular solution of 6·2 (2), satisfying the condition that the temperature and its differential coefficient are continuous across the boundary $x = H$, is

$$V_0 = A\{H^2 - (x-H)^2\}/2k \text{ when } x \text{ is less than } H \quad \text{.........(1)},$$

$$V_0 = AH^2/2k \text{ when } x \text{ is greater than } H \quad \text{.........(2)}.$$

We are given that when t is zero

$$V = S + mx \quad \text{.........(3)}.$$

Hence

$$f(x) = S + mx - A\{H^2 - (x-H)^2\}/2k \text{ when } x \text{ is less than } H \text{ ...(4)},$$

$$f(x) = S + mx - AH^2/2k \text{ when } x \text{ is greater than } H \quad \text{.........(5)}.$$

On substituting from these into 6·2 (4) we find the value of V_1, by integration, to be

$$\begin{aligned} V_1 = {} & S \operatorname{Erf} \frac{x}{2ht^{\frac{1}{2}}} + mx \\ & + \frac{A}{2k}\Big[(x^2 + 2h^2t)\operatorname{Erf}\frac{x}{2ht^{\frac{1}{2}}} - \tfrac{1}{2}\{(x-H)^2 + 2h^2t\}\operatorname{Erf}\frac{x-H}{2ht^{\frac{1}{2}}} \\ & - \tfrac{1}{2}\{(x+H)^2 + 2h^2t\}\operatorname{Erf}\frac{x+H}{2ht^{\frac{1}{2}}} + 2xht^{\frac{1}{2}}\pi^{-\frac{1}{2}}e^{-x^2/4h^2t} \\ & - ht^{\frac{1}{2}}\pi^{-\frac{1}{2}}(x-H)e^{-(x-H)^2/4h^2t} - ht^{\frac{1}{2}}\pi^{-\frac{1}{2}}(x+H)e^{-(x+H)^2/4h^2t}\Big] \quad \text{...(6)}. \end{aligned}$$

It may be verified by direct differentiation that this expression satisfies all the conditions.

6·31. If we differentiate V with regard to x and then put x equal to zero, we obtain the temperature lapse rate at the surface. Thus

$$\left(\frac{\partial V}{\partial x}\right)_{x=0} = m + \frac{S}{h\sqrt{(\pi t)}} + \frac{A}{k}\left[H - H\operatorname{Erf}\frac{H}{2h\sqrt{t}} + h\sqrt{\frac{t}{\pi}}\left(1 - e^{-H^2/4h^2t}\right)\right] \quad \text{.........(1)}.$$

The present value of $\left(\frac{\partial V}{\partial x}\right)_0$ is $0°{\cdot}00032$ C. per cm. We have already seen in 6·11 that m can hardly exceed $0°{\cdot}000025$ C. per cm. S is the melting point

of average continental rocks. This is somewhat uncertain, for two reasons. The cooling has extended to a considerable depth, and therefore the depth that the thermal constants used should refer to is doubtful. Indeed it is certain that they must vary somewhat with depth, and therefore the above solution, which depends essentially on their constancy, cannot be quite accurate. The melting points of basalts and granites are from 1060° to 1240°* (although some attention will have to be devoted later to the meaning of the melting point of mixed rocks such as these). S will here be supposed equal to 1200°. The time since the solidification of the earth has already been stated to be greater than the age of the oldest known rocks, namely about $1{\cdot}4 \times 10^9$ years, but probably the interval between the formation of the earth and the condensation of the ocean was not a large fraction of this. We have no means of ascertaining the length of the interval between the formation of the ocean and of the oldest known igneous rocks. These are, however, of very old geological date, and it is at least plausible that this interval also is not a large fraction of the time since the formation of the rocks in question. Accordingly it will be supposed here that the time since the solidification of the earth is $1{\cdot}6 \times 10^9$ years, or 5×10^{16} seconds.

6·32. For most acid crustal rocks k is about 0·006 c.g.s. For basic rocks it is somewhat lower, about 0·004. Both types of rocks will be involved in the present problem, and accordingly the value adopted will be 0·005 c.g.s. The density being taken as 2·8 gm./cm.^{3}, and the specific heat as 0·25 cal./gm. 1° C., we have

$$h = 0{\cdot}084 \qquad \text{.........(1).}$$

We already have for surface rocks

$$A = 1{\cdot}0 \times 10^{-12} \qquad \text{.........(2).}$$

Thus every quantity involved in equation 6·31 (1) is known except H. This equation may now be solved to find H. If we put

$$H/2ht^{\frac{1}{2}} = l \qquad \text{.........(3),}$$

this equation becomes

$$l\,(1 - \operatorname{Erf} l) + \frac{1}{2\sqrt{\pi}}(1 - e^{-l^2}) = 0{\cdot}034 \qquad \text{.........(4),}$$

giving

$$l = 0{\cdot}035 \qquad \text{.........(5),}$$

and

$$H = 1{\cdot}31 \times 10^6 \text{ cm.} = 13 \text{ km.} \qquad \text{.........(6).}$$

Thus the depth of the radioactive layer is 13 km. We notice incidentally that out of the observed value of $\left(\frac{\partial V}{\partial x}\right)_0$, less than a tenth is provided by the term in m, while the term involving S contributes $0^\circ{\cdot}00004$/cm. The remaining 0·00025 comes from radioactive sources, and practically the whole of this from the term AH/k, which is independent of the time.

* F. W. Clarke, *Data of Geochemistry*, 1916, 296.

Thus only about 13 per cent. of the heat being conducted out of the earth at the present time is due to cooling of the interior; the rest is supplied by radioactivity, and will continue until the latter fails through exhaustion of the uranium and thorium in the crust. We notice further that each term in (1) is positive, and therefore if H exceeded 16 km. it would be impossible for the quantity on the right to be as small as that on the left, even if no allowance whatever was made for initial heat and for the initial rate of increase of temperature inwards. Thus the concentration of radioactivity in the surface layers is a necessary consequence of our assumptions.

6·4. *Radioactivity Decreasing with Depth.* If instead of supposing radioactivity uniform down to a certain depth, we assume it to fall off exponentially with the depth, we shall have

$$P = Ae^{-ax} \qquad \text{.........(1)},$$

where a is a constant and A is the same as before. In these circumstances we have

$$V_0 = \frac{A}{a^2k}(1 - e^{-ax}) \qquad \text{.........(2)},$$

$$f(x) = S + mx - \frac{A}{a^2k}(1 - e^{-ax}) \qquad \text{.........(3)},$$

$$V = \left(S - \frac{A}{a^2k}\right)\operatorname{Erf}\lambda + mx + \frac{A}{a^2k}(1 - e^{-ax})$$
$$+ \frac{A}{2a^2k}e^{\gamma^2}[e^{-ax}\{1 - \operatorname{Erf}(\gamma - \lambda)\} - e^{ax}\{1 - \operatorname{Erf}(\gamma + \lambda)\}] \qquad \text{.........(4)},$$

where

$$\lambda = x/2ht^{\frac{1}{2}} \qquad \text{.........(5)},$$

$$\gamma = aht^{\frac{1}{2}} \qquad \text{.........(6)}.$$

As before, it may be verified by differentiation that this expression satisfies the conditions. Differentiating with regard to x, and then putting x zero, we have

$$\left(\frac{\partial V}{\partial x}\right)_{x=0} = m + \frac{S}{h\sqrt{\pi t}} + \frac{A}{ak}\left(1 - \frac{1}{ah\sqrt{\pi t}}\right) \qquad \text{.........(7)}.$$

Just as before, all the quantities in this equation are known except one, in this case a, for which we can therefore solve.

With the values already adopted for the quantities involved, we have

$$\frac{1}{a} = 1{\cdot}30 \times 10^6 \text{ cm.} = 13 \text{ km.} \qquad \text{.........(8)}.$$

This is practically the same as the value of H found on the hypothesis of a uniformly radioactive surface layer; a consequence evidently implied by the fact that equations 6·31 (1) and 6·4 (7) differ only in their small terms. On either hypothesis it is seen that radioactivity must fall off rapidly with depth, and probably become insignificant at something of the order of 50 km. below the surface. If this were not so, the rate of increase of

temperature downwards would necessarily be greater than is observed, and it would therefore be impossible to coordinate our data.

6·5. *Hypotheses to account for the Distribution of Radioactivity.* The falling off of radioactivity with depth is capable of two interpretations. It may be suggested that radioactive elements actually occur at great depths, but that their disintegration is prevented by the pressures acting on them. This hypothesis is scarcely plausible. In experiments already made* it has been shown that radioactivity is not appreciably affected by pressures up to $2{\cdot}6 \times 10^{10}$ dynes/cm.^{2}, or by temperatures up to 2500° C. The latter temperature is greater than can occur in the crust on the theory here presented, and the pressure corresponds to a depth of about 100 km. in the crust. Thus the reduction in radioactivity just shown to be probable at a depth of 13 km. would require that the pressure at that depth had an inhibitory effect greater than is produced by any pressure less than or equal to that at 100 km. Hence this interpretation of the diminution is unsatisfactory.

6·51. The alternative is that there is a real diminution with depth in the quantity of radioactive substances present. It has already been seen that they are more abundant in acid than in basic rocks, and considerations of density and direct geological evidence both indicate that basicity should increase with depth. The latter considerations, however, give no clue as to the amount of acid rock in the crust, and therefore cannot determine quantitatively how basicity varies with depth. The theory just developed, however, gives a clue. The radioactivity of average basic rocks is rather less than $\frac{1}{3}$ of that of average acid rocks, and on the exponential law of decrease such an intensity would correspond to a depth of $13 \log \dfrac{1{\cdot}01}{0{\cdot}29}$ km., or about 16 km. Hence the theory enables us to make the quantitative suggestion that basic rocks are predominant in the crust at a depth not greater than 16 km. This suggestion will be seen later (12·8) to have some interest in connection with the structure of the continents.

6·52. The explanation of the concentration of radioactive materials in the acid rocks is still somewhat uncertain. It has often been thought that the high densities of these elements would tend to make them collect towards the centre of the earth, contrary to what has just been inferred. Holmes hints† that the concentrating action of volatile fluxes is important in this connection. The radioactive elements form volatile compounds, which would be expected to accumulate towards the top in a fluid earth, and thus would become concentrated in the granitic layer.

6·6. *Cooling at Great Depths.* Let us next consider the temperature distribution at great depths. It has been seen that $H/2ht^{\frac{1}{2}}$ and $1/2aht^{\frac{1}{2}}$

* Eve and Adams, *Nature*, July 1907, 209.

† *Geological Magazine*, Feb. 1915, 64.

are both small numbers, and therefore we may expand the temperature in powers of them. If q is moderate and l small, we have to the second order in l

$$\operatorname{Erf}(q+l) = \operatorname{Erf} q + \frac{2l}{\sqrt{\pi}} e^{-q^2} - \frac{2l^2}{\sqrt{\pi}} q e^{-q^2} \qquad \text{.........(1)},$$

$$e^{-(q+l)^2} = e^{-q^2}(1 - 2ql + 2q^2l^2 - l^2) \qquad \text{.........(2)}$$

Considering first the uniform distribution of radioactive matter, we may approximate to 6·3 (6) by means of (1) and (2), and find that V reduces to the form

$$V = mx + \left(S - \frac{AH^2}{2k}\right) \operatorname{Erf} \frac{x}{2ht^{\frac{1}{2}}} + \frac{AH^2}{2k} - \mu \qquad \text{.........(3)},$$

where $\qquad \mu = 0$ if x is greater than H $\qquad$ ....(4),

and $\qquad \mu = A(H - x)^2/2k$ if x is less than H $\qquad$ ....(5)

Finally, if V' denote the change in temperature at any depth since solidification, we have

$$V' = -\left(S - \frac{AH^2}{2k}\right)\left(1 - \operatorname{Erf} \frac{x}{2ht^{\frac{1}{2}}}\right) \qquad \text{.........(6)}.$$

With the data adopted we have $2ht^{\frac{1}{2}} = 370$ km.; while $AH^2/2k$ is about 170°, and therefore $S - \dfrac{AH^2}{2k}$ will be about 1030°. At a depth of 370 km. therefore, the cooling since solidification will be 162°. At 700 km it will be 8°. Thus considerable cooling will have taken place at depths comparable with 300 or 400 km., but not at double these depths. Thus the matter at depths of 700 km. or more must be in almost the same state as just after the earth first solidified. Incidentally it will be noticed that, on account of the smallness of $AH^2/2k$ in comparison with S, the cooling at these great depths must be almost the same as if the earth had been cooling for the actual time that has elapsed since its solidification, without cooling having been in the least delayed by radioactivity.

6·61. Considering next the exponential distribution, we find that $\gamma = 12$, and therefore whatever value of λ less than about 10 we consider, we can approximate to $\operatorname{Erf}(\gamma - \lambda)$ and $\operatorname{Erf}(\gamma + \lambda)$, and the approximate form of the temperature is

$$V = mx + \left(S - \frac{A}{a^2k}\right) \operatorname{Erf} \lambda + \frac{A}{a^2k}(1 - e^{-ax}) + \frac{A\lambda e^{-\lambda^2}}{a^2k\sqrt{\pi}(\gamma^2 - \lambda^2)} \qquad \text{......(1)}.$$

With the numerical values already obtained, the last term is at most of order $1/\gamma^2$, in comparison with the third, and γ is about 12. The second term, again, is greater than the third. Hence, the last term may be neglected, in comparison with the second and third.

The change of temperature at any depth since solidification is

$$V' = -\left(S - \frac{A}{a^2k}\right)(1 - \operatorname{Erf} \lambda).$$

If a is 13 km., we find $A/a^2k = 340°$, and therefore $S - A/a^2k$ is 860°. The cooling at a depth of 370 km. is 135°, and that at 700 km. is 7°. Thus the

cooling at great depths is about 83 per cent. of that found on the hypothesis of a uniform distribution of radioactive matter down to a certain depth.

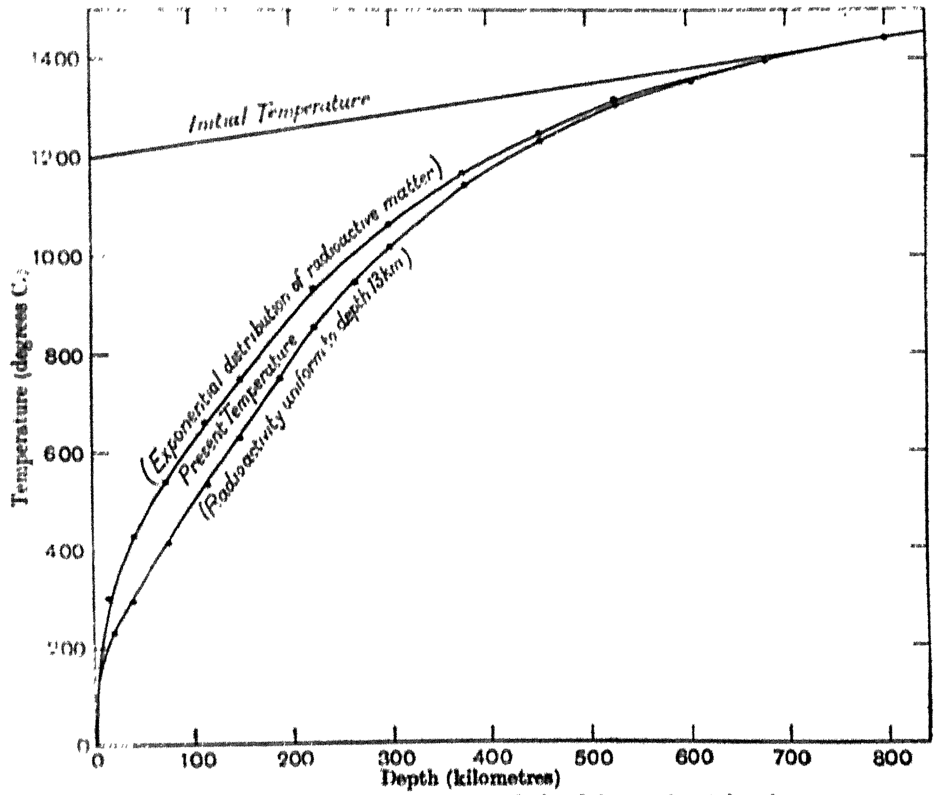


Fig. 4. Distribution of temperature with depth in continental regions.

6·7. *Effect of a Lower Melting Point.* It has been suggested that the melting point of rocks so far adopted is too high, and that the actual melting point is considerably lowered by the presence of volatile constituents, and may be as low as 800°. With this hypothesis H and $1/a$ are both increased by about 5 per cent. Thus $S - AH^2/2k$ is reduced to about 610°, and the corresponding factor $S - A/a^2k$ that occurs in the theory of the exponential distribution is only 470°. Thus the cooling at all depths is considerably reduced if this hypothesis is correct. For oceanic regions, with the same melting point, we have

$$A/a^2k = 110° \text{ and } S - A/a^2k = 690°.$$

6·8. *The Cooling of Suboceanic Rocks.* The discussion has so far referred to continental rocks alone. It rests ultimately on the observations of the increase of temperature inwards in the earth's crust, made in mines and borings, and up to the present no mine or boring has been made in the ocean bottom. If such an undertaking could be carried out, most valuable information could be obtained; but at present the cooling of the suboceanic

crust can only be inferred by means of a fairly plausible hypothesis. There is some reason to believe that the surface rocks of the ocean bed are predominantly basic. This has been inferred from several facts, especially the greater basicity of volcanic rocks from oceanic islands. It will be assumed here that the surface rocks have a content of radioactive matter equal to that of basic igneous rocks within the continents, giving a rate of heat production of $0{\cdot}29 \times 10^{-12}$ cal./cm.^{3} sec. The rate of evolution of heat will be supposed to decrease according to the exponential law, with the same value of a as in the case of the continents. The problem of the cooling of the regions below the ocean becomes determinate with these assumptions, which would hold accurately if the ocean had been made by the removal of a uniform thickness of matter from the continental surface; this process bears a slight resemblance to what appears to be required by any theory of continental formation. We find that A/a^2k is $100°$, so that $S = A/a^2k$ is $1100°$. The cooling at a depth of 370 km. is about $180°$, and that at 700 km. is $10°$. The cooling below the oceans is therefore substantially greater than that below the continents.

6·9. *Summary.* The theory of the cooling of the earth developed in this chapter has made it possible to find quantitatively the amount of cooling that has taken place at any depth since solidification, either below the continents or below the oceans. This theory has been definitely based on certain simplifying assumptions; in particular it has been supposed that the conductivity and other thermal properties of the rocks affected are constant. We have no direct evidence concerning the properties of rocks at depths exceeding a few kilometres. The assumptions of constancy therefore amount virtually to the assumption that these properties are unaffected by depth, or in other words by changes of pressure and temperature. Direct experiment in the laboratory has suggested that the effects of such changes are insufficient to require any serious modification of the results obtained. The evidence, however, remains inconclusive, and to this extent the theory is doubtful. This fact, however, is no reason for neglecting the theory. In the absence of direct evidence that there is a physical connection between two quantities, the most probable hypothesis is always the simplest, namely, that they are independent*; and the scientific procedure is to push ahead as far as possible with the most probable hypothesis on the data in our possession. If then the hypothesis is found to lead to incorrect consequences, that will be additional evidence, and the hypothesis will have to be replaced by a new one that becomes more probable when the new data are taken into account; but even in that case the working out of the hypothesis of independence in a form capable of being tested will have been a necessary stage in the acquirement of the new knowledge, and will therefore have been of scientific value.

* Wrinch and Jeffreys, *Phil. Mag.* **45**, 1923, 368–74.

So long as the hypothesis, on the other hand, leads to results in harmony with our knowledge, it remains the most probable on the data we have, and its consequences are scientific inferences in the usual sense of the term.

6·91. The results are, therefore, that there is good reason for suspecting that the rocks in continental regions at a depth of 16 km. are typically basic, although it is known that the typical rocks of the crust are acidic. Further, the cooling of the earth since solidification amounts, at a depth of some 300 km., to between 200° and 300°; at a depth of 700 km. the cooling is as yet insignificant. Under the oceans the cooling must be somewhat greater at all depths. The physical significance of these results is that the physical state of the matter at great depths can scarcely have changed since solidification, and therefore it may be expected to show in some respects at least the usual properties of a liquid; namely, it should be capable of being deformed to any extent by a shearing stress, however small, provided this acted for a long enough time; and it should be devoid of rigidity, and therefore unable to transmit distortional waves. We shall return later to the consideration of how far these predictions are actually satisfied. At the moderate depth of 200 to 400 km., it would be expected that the tendency to fluidity would already be beginning to show itself, probably by softening, which usually commences at temperatures far below the melting point when an impure substance is heated. Thus rocks at such depths would yield slowly to small shearing stresses; but it is possible that the stress required to produce permanent deformation must exceed a definite limit, even though this limit must be much less than in the same substances at ordinary temperatures.

CHAPTER VII

The Equations of Motion of an Elastic Solid with Initial Stress

"Lasciate ogni speranza, voi ch' entrate." DANTE, *Inferno*

7·1. In the following pages several results in the theory of elasticity will be utilized, and it will be convenient to discuss them as far as possible consecutively. If a stress is applied to any substance but a perfect fluid, and we consider a small element of surface within the substance, the stress across the element (i.e. the force per unit area exerted by the matter on one side of it upon the matter on the other side) has three components, one of which is normal to the element, while the other two are tangential to the element. In a perfect fluid the two tangential components are zero.

Consider then a small parallelepiped within the substance, its centre being at the point (x, y, z), and its sides being parallel to the coordinate axes and of lengths dx, dy, dz. Let us denote the components of force per unit area across a plane perpendicular to the axis of x by p_{xx}, p_{xy}, p_{xz}. By considering the opposite faces of the parallelepiped in pairs, we easily see that the resultant force on the element due to the stresses has components

$$\left(\frac{\partial p_{xx}}{\partial x} + \frac{\partial p_{yx}}{\partial y} + \frac{\partial p_{zx}}{\partial z}\right) dx\,dy\,dz \qquad \ldots\ldots (1),$$

and two symmetrical expressions. If ρ is the density and (X, Y, Z) the bodily force per unit mass, the force on the element arising from this is

$$(\rho X, \rho Y, \rho Z)\, dx\,dy\,dz \qquad \ldots\ldots (2)$$

The two sets of forces together produce the acceleration of the element. If u, v, w denote its displacements from its initial position we shall therefore have

$$\rho \frac{d^2u}{dt^2} = \rho X + \frac{\partial p_{xx}}{\partial x} + \frac{\partial p_{yx}}{\partial y} + \frac{\partial p_{zx}}{\partial z} \qquad \ldots\ldots (3),$$

with two similar equations. In the differentiation with regard to the time the displacement is supposed to have been expressed as a function of the initial coordinates and the time, and the differentiation is a partial one with regard to the time in this system, so as to give the true acceleration of each element.

Now let us consider how the system of stresses $p_{xx}, p_{xy}, p_{xz}, \ldots p_{zz}$ can arise. Suppose that in the initial state they had the values indicated by the index zero, $p^0_{xx}, p^0_{xy}, \ldots p^0_{zz}$. When an element is displaced the stresses on it in general change, and corresponding changes in its size and form occur. Suppose, for the sake of generality, that at the same time an increase of temperature V takes place, and that the coefficient

of linear expansion is n. Then the alterations in dimensions due to simple expansion in the absence of stress would be such as to make

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = \frac{\partial w}{\partial z} = nV; \ \frac{\partial u}{\partial y} = \frac{\partial u}{\partial z} = \frac{\partial v}{\partial x} = \frac{\partial v}{\partial z} = \frac{\partial w}{\partial x} = \frac{\partial w}{\partial y} = 0 \ \ldots(4).$$

In the actual displacement, however, the values of the displacements will not be such as to satisfy these conditions; the size and form of the element will change on account of the changes in the stresses from neighbouring elements. Now it is proved in works on elasticity that the deformation produced by a small extra stress is related to it according to the laws

$$p'_{xx} = \lambda\,(e_{xx} + e_{yy} + e_{zz}) + 2\mu e_{xx}, \text{ etc.} \qquad \ldots\ldots\ldots(5),$$

$$p'_{xy} = p'_{yx} = \mu e_{xy} = \mu e_{yx} \qquad \ldots\ldots\ldots(6),$$

where accents denote that we are dealing with the changes in the stresses. Here $e_{xx}, e_{xy}, e_{xz}, \ldots e_{zz}$ are the changes in the quantities

$$\frac{\partial u}{\partial x}, \ \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}, \ \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}, \ \ldots, \ \frac{\partial w}{\partial z},$$

measured from the state before the deforming stresses are applied, and λ and μ are the two elastic constants of the element, supposed isotropic. The additional stresses are supposed small enough for their squares to be neglected. Thus we shall have

$$e_{xx} = \frac{\partial u}{\partial x} - nV, \text{ etc.} \qquad \ldots\ldots\ldots(7),$$

$$e_{xy} = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}, \text{ etc.} \qquad \ldots\ldots\ldots(8),$$

and

$$p'_{xx} = \lambda\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) + 2\mu\frac{\partial u}{\partial x} - (3\lambda + 2\mu)\,nV, \text{ etc.} \ \ldots\ldots(9),$$

$$p'_{xy} = \mu\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) \qquad \ldots\ldots\ldots(10).$$

Thus we have for the equations of motion

$$\begin{aligned}\rho\frac{d^2u}{dt^2} = \rho X &+ \frac{\partial p^0_{xx}}{\partial x} + \frac{\partial p^0_{yx}}{\partial y} + \frac{\partial p^0_{zx}}{\partial z} \\ &+ \frac{\partial}{\partial x}\left(\lambda\delta + 2\mu\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left\{\mu\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)\right\} + \frac{\partial}{\partial z}\left\{\mu\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)\right\} \\ &- \frac{\partial}{\partial x}\{(3\lambda + 2\mu)\,nV\} \qquad \ldots\ldots\ldots(11),\end{aligned}$$

with two symmetrical equations, where δ has been written for

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \qquad \ldots\ldots\ldots(12).$$

We may also write γ for $\quad (3\lambda + 2\mu)\,nV \qquad \ldots\ldots\ldots(13).$

It will be noticed that the bodily force (X, Y, Z) is to be considered as evaluated at (x, y, z) after the deformation. Let (X_0, Y_0, Z_0) be the force at (x, y, z) before the deformation, and put

$$X = X_0 + X_1, \text{ etc.} \qquad \ldots\ldots\ldots(14).$$

The stress components p^0_{xx}, p^0_{xy}, ... p^0_{zz}, on the other hand, refer to the position before deformation of the particle that is brought to (x, y, z) by the deformation. If we denote by q_{xx}, q_{xy}, ... q_{zz} the corresponding stresses before deformation on the element that was at (x, y, z) before deformation, we shall have for small displacements

$$p^0_{xx} = q_{xx} - u\frac{\partial q_{xx}}{\partial x} - v\frac{\partial q_{xy}}{\partial y} - w\frac{\partial q_{xz}}{\partial z} \qquad \text{.........(15)},$$

with corresponding relations for the other components. Also let ρ_0 be the density at (x, y, z) before deformation, and put

$$\rho = \rho_0 + \rho_1 \qquad \text{.........(16)}.$$

Then

$$\rho = \rho_0 - \frac{\partial}{\partial x}(\rho_0 u) - \frac{\partial}{\partial y}(\rho_0 v) - \frac{\partial}{\partial z}(\rho_0 w) \qquad \text{.........(17)}$$

is the equation of continuity, and determines ρ_1. If the initial state was one of equilibrium, the equations of motion must hold when u, v, and w are put equal to zero. Thus we have

$$0 = \rho_0 X_0 + \frac{\partial q_{xx}}{\partial x} + \frac{\partial q_{yx}}{\partial y} + \frac{\partial q_{zx}}{\partial z} \qquad \text{.........(18)},$$

with two symmetrical relations. On subtracting (18) from (11), and omitting terms involving squares and products of the displacements, we have

$$\begin{aligned}\rho_0\frac{d^2u}{dt^2} = \rho_0 X_1 + \rho_1 X_0 &- \frac{\partial}{\partial x}\left(u\frac{\partial q_{xx}}{\partial x} + v\frac{\partial q_{xx}}{\partial y} + w\frac{\partial q_{xx}}{\partial z}\right)\\ &- \frac{\partial}{\partial y}\left(u\frac{\partial q_{xy}}{\partial x} + v\frac{\partial q_{xy}}{\partial y} + w\frac{\partial q_{xy}}{\partial z}\right) - \frac{\partial}{\partial z}\left(u\frac{\partial q_{xz}}{\partial x} + v\frac{\partial q_{xz}}{\partial y} + w\frac{\partial q_{xz}}{\partial z}\right)\\ &+ \frac{\partial}{\partial x}\left(\lambda\delta + 2\mu\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left\{\mu\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)\right\} + \frac{\partial}{\partial z}\left\{\mu\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)\right\} - \frac{\partial\gamma}{\partial x}\end{aligned} \qquad \text{.........(19)}.$$

Further progress requires a knowledge of the initial stress. If the substance was under no shearing stress before the displacement, as would be correct if complete yield had taken place to any stresses produced during solidification, we should have

$$q_{xy} = q_{yx} = q_{xz} = q_{zx} = q_{yz} = q_{zy} = 0 \qquad \text{.........(20)},$$

$$\rho_0 X_0 + \frac{\partial q_{xx}}{\partial x} = 0, \quad q_{xx} = q_{yy} = q_{zz} = -p, \text{ say} \qquad \text{.........(21)}.$$

With this hypothesis, (19) reduces to

$$\begin{aligned}\rho_0\frac{d^2u}{dt^2} &= \rho_0 X_1 + \rho_1 X_0 + \frac{\partial}{\partial x}\{\rho_0(uX_0 + vY_0 + wZ_0)\}\\ &+ \frac{\partial}{\partial x}\left(\lambda\delta + 2\mu\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left\{\mu\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)\right\} + \frac{\partial}{\partial z}\left\{\mu\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)\right\} - \frac{\partial\gamma}{\partial x}\end{aligned} \qquad \text{......(22)}.$$

CHAPTER VIII

The Bending of the Earth's Crust by the Weight of Mountains

"A pupil began to learn geometry with Euclid and asked, when he had learnt one proposition, 'What advantage shall I get by learning these things?' And Euclid called the slave and said, 'Give him sixpence, since he must needs gain by what he learns.'" Sir T. L. HEATH, *A History of Greek Mathematics*.

8·1. In what follows a great deal of attention will be devoted to the mechanism that enables the earth's crust to support the weight of mountains without immediately yielding and gradually effacing all departures from perfect symmetry. A discussion of the stresses in the crust produced by the weight of surface inequalities will therefore be necessary. For this purpose the crust will be supposed to be of finite thickness, and a heavy fluid will be supposed to lie below it. Thus the effect of any additional pressure over the upper surface will be to bend the crust downwards, and the under surface will be forced down into the fluid against hydrostatic pressure. The final deformation will therefore be the result of the joint action of the additional pressures on the upper and lower boundaries.

The problem is a statical one; the accelerations in the equations of motion are zero. There is no change of temperature, so that γ is zero. The elasticity terms in the equations of motion are of order

$$(\lambda + 2\mu)\frac{\partial^2 w}{\partial x^2} \text{ or } (\lambda + 2\mu)\frac{\partial^2 w}{\partial x \partial z},$$

while the gravity terms are comparable with

$$\frac{\partial}{\partial x}(\rho_0 wg) \text{ or } \frac{\partial}{\partial z}(\rho_0 wg).$$

If the wave length of the disturbing pressure is $2\pi/\kappa$, gravity will be unimportant in comparison with elasticity if $g\rho_0$ is small compared with $(\lambda + 2\mu)\kappa$. Now the velocities of the longitudinal waves of earthquakes near the surface enable us to say that $(\lambda + 2\mu)/\rho_0$ is about 5×10^{11} cm.^{2}/sec.^{2} Thus our condition is that κ shall be large compared with 2×10^{-9}/cm., or that $2\pi/\kappa$ shall be small compared with 30,000 km. The effect of bodily gravity will in fact be small in comparison with that of elasticity unless the stresses are of a wave length comparable with the circumference of the earth, when it would be necessary to include curvature in our calculations as well. Thus the terms depending on bodily forces can be neglected when we are dealing with the support of mountain ranges; the curvature of the earth can be neglected in similar problems; and it will therefore be legitimate to treat the problem as one of simple stress in a uniform flat plate of infinite horizontal extent.

8·2. Let us take the axis of z vertically downwards, and those of x and y horizontal. The origin will be in the middle of the layer, and the

thickness of the layer will be $2h$, so that the upper and lower surfaces in their unstrained state are $z = -h$ and $z = h$ respectively. The layer is supposed homogeneous. The equations 7·1 (22) then reduce to

$$(\lambda+\mu)\frac{\partial\delta}{\partial x} + \mu\nabla^2 u = 0 \qquad \text{.........(1),}$$

$$(\lambda+\mu)\frac{\partial\delta}{\partial y} + \mu\nabla^2 v = 0 \qquad \text{.........(2),}$$

$$(\lambda+\mu)\frac{\partial\delta}{\partial z} + \mu\nabla^2 w = 0 \qquad \text{.........(3).}$$

By differentiating with respect to x, y, and z respectively and adding, we find

$$\nabla^2\delta = 0 \qquad \text{.........(4).}$$

Let us write p for $\frac{\partial}{\partial x}$, q for $\frac{\partial}{\partial y}$, and ϑ for $\frac{\partial}{\partial z}$. Put also

$$p^2 + q^2 = -r^2 \qquad \text{.........(5).}$$

Then
$$\nabla^2 = \vartheta^2 + p^2 + q^2 = \vartheta^2 - r^2 \qquad \text{.........(6).}$$

Solving for δ by the symbolical method of Heaviside and Bromwich, r being independent of z, we have

$$\delta = A\cosh rz + B\sinh rz \qquad \text{.........(7),}$$

where A and B are functions of x and y only. Let us now introduce a function χ such that

$$\chi = \frac{A}{2r} z\sinh rz + \frac{B}{2r} z\cosh rz \qquad \text{.........(8).}$$

Then we readily verify by differentiation that

$$(\vartheta^2 - r^2)\chi = \delta \qquad \text{.........(9).}$$

If we substitute for δ in (1), (2) and (3), we have three equations to find u, v, and w. Particular integrals will evidently be

$$(u_0, v_0, w_0) = -\frac{\lambda+\mu}{\mu}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\chi \qquad \text{.........(10),}$$

since with these values

$$(\lambda+\mu)\frac{\partial\delta}{\partial x} + \mu\nabla^2 u = (\lambda+\mu)\frac{\partial\delta}{\partial x} - \mu(\vartheta^2 - r^2)\frac{\lambda+\mu}{\mu}\frac{\partial\chi}{\partial x} \quad \text{...(11),}$$

which is zero by (9). The complementary functions are any solutions of

$$(\vartheta^2 - r^2)\phi = 0.$$

We have therefore

$$(u, v, w) = -\frac{\lambda+\mu}{\mu}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\chi + (u_1, v_1, w_1) \quad \text{......(12),}$$

where we can write
$$u_1 = U_0\cosh rz + U_1\sinh rz \qquad \text{.........(13),}$$

$$v_1 = V_0\cosh rz + V_1\sinh rz \qquad \text{.........(14),}$$

$$w_1 = W_0\cosh rz + W_1\sinh rz \qquad \text{.........(15).}$$

Here U_0, U_1, V_0, V_1, W_0, W_1 are functions of x and y alone. We still have the condition that
$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=\delta \qquad \text{.........(16).}$$
Substituting in this from (12), we have
$$pu_1+qv_1+\vartheta w_1=\frac{\lambda+2\mu}{\mu}\delta \qquad \text{.........(17).}$$
This must hold for all values of z; so equating coefficients of $\cosh rz$ and $\sinh rz$ we have
$$pU_0+qV_0+rW_1=\frac{\lambda+2\mu}{\mu}A \qquad \text{.........(18),}$$
$$pU_1+qV_1+rW_0=\frac{\lambda+2\mu}{\mu}B \qquad \text{.........(19).}$$

8·21. We have still to make use of the boundary conditions. These will be supposed to be that there is no shearing stress across the planes $z=\pm h$, and that

when $z=-h$, $\qquad p_{zz}=P+Q$ ....(20),

when $z=h$, $\qquad p_{zz}=P-Q$ ....(21).

We have from 7·1 (10)
$$p_{zx}=\mu(pw+\vartheta u)=-2(\lambda+\mu)p\vartheta\chi+\mu(pw_1+\vartheta u_1)$$
$$=-(\lambda+\mu)\frac{p}{r}(A\sinh rz+B\cosh rz)-(\lambda+\mu)pz(A\cosh rz+B\sinh rz)$$
$$+\mu\{p(W_0\cosh rz+W_1\sinh rz)+r(U_0\sinh rz+U_1\cosh rz)\} \text{......(22).}$$
The vanishing of this where $z=\pm h$ gives the two relations
$$-(\lambda+\mu)\frac{pA}{r}(1+rh\coth rh)+\mu(pW_1+rU_0)=0 \text{......(23),}$$
$$-(\lambda+\mu)\frac{pB}{r}(1+rh\tanh rh)+\mu(pW_0+rU_1)=0 \text{......(24).}$$
The first of these can be written
$$W_1+\frac{r}{p}U_0=\frac{\lambda+\mu}{\mu r}(1+rh\coth rh)A \qquad \text{.........(25),}$$
and we see by symmetry that the vanishing of p_{yz} at the boundaries will prove the right side of (25) equal to $W_1+\frac{r}{q}V_0$. Similarly (24) gives
$$W_0+\frac{r}{p}U_1=\frac{\lambda+\mu}{\mu r}(1+rh\tanh rh)B \qquad \text{.........(26)}$$
$$=W_0+\frac{r}{q}V_1.$$
Using these relations to eliminate U_0, V_0, U_1, V_1 from (18) and (19), we find
$$W_1=-\frac{\lambda+2\mu}{2\mu r}A+\frac{\lambda+\mu}{2\mu r}(1+rh\coth rh)A \qquad \text{.........(27),}$$
$$W_0=-\frac{\lambda+2\mu}{2\mu r}B+\frac{\lambda+\mu}{2\mu r}(1+rh\tanh rh)B \qquad \text{.........(28).}$$

We have also
$$p_{zz} = \lambda\delta + 2\mu \vartheta w$$
$$= \lambda\delta - 2(\lambda+\mu)\vartheta^2\chi + 2\mu\vartheta w_1 \quad \text{.........(29)}.$$

But
$$\vartheta^2\chi = r^2\chi + \delta.$$

Hence
$$p_{zz} = -(\lambda + 2\mu)\delta - 2(\lambda+\mu) r^2\chi + 2\mu\vartheta w,$$
$$= -(\lambda+2\mu)(A\cosh rz + B\sinh rz)$$
$$-(\lambda+\mu) rz (A\sinh rz + B\cosh rz)$$
$$+ 2\mu r (W_0 \sinh rz + W_1 \cosh rz) \quad \text{.........(30)}.$$

By hypothesis p_{zz} is to be equal to $P+Q$ when $z = -h$, and to $P - Q$ when $z = h$. Hence

$$-\{(\lambda+2\mu) + (\lambda+\mu) rh\tanh rh\} A + 2\mu r W_1 = P\operatorname{sech} rh \quad \text{...(31)},$$
$$-\{(\lambda+2\mu) + (\lambda+\mu) rh\coth rh\} B + 2\mu r W_0 = -Q\operatorname{cosech} rh \text{...(32)}.$$

Substituting for W_0 and W_1 from (27) and (28), we have

$$(\lambda+\mu)\{1 + rh\operatorname{cosech} rh\operatorname{sech} rh\} A = P\operatorname{sech} rh \quad \text{......(33)},$$
$$(\lambda+\mu)\{1 - rh\operatorname{cosech} rh\operatorname{sech} rh\} B = -Q\operatorname{cosech} rh \text{...(34)}.$$

These equations, with the expressions for the other coefficients in terms of A and B, give the symbolic solution of the problem when P and Q are known. In particular, if P and Q are proportional to $\sin\kappa x \sin\kappa' y$, where κ and κ' are constants, or to any other product of harmonic functions of κx and $\kappa' y$, the solution is to be found by simply substituting $(\kappa^2 + \kappa'^2)^{\frac{1}{2}}$ for r.

8·22. We require to know the stress-differences in the solid layer. The stresses are given by

$$p_{xx} = \lambda\delta + 2\mu\frac{\partial u}{\partial x}, \text{ with two similar expressions;}$$
$$p_{xy} = \mu\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right), \text{ with two similar expressions.}$$

Substituting for the coefficients in these, we find

$$p_{zz} = (\lambda+\mu)\{(1 + rh\coth rh) A\cosh rz + (1 + rh\tanh rh) B\sinh rz$$
$$- rz(A\sinh rz + B\cosh rz)\}\text{.........(35)},$$

$$p_{xx} = A\cosh rz\left(\lambda - \mu\frac{p^2}{r^2} + (\lambda+\mu)\frac{p^2}{r} h\coth rh\right) + B\sinh rz\left(\lambda - \mu\frac{p^2}{r^2}\right.$$
$$\left. + \frac{(\lambda+\mu)p^2}{r} h\tanh rh\right) - \frac{(\lambda+\mu)p^2}{r} z(A\sinh rz + B\cosh rz)\text{...(36)},$$

$$p_{zx} = (\lambda+\mu)\frac{p}{r}\{Arh\coth rh\sinh rz + Brh\tanh rh\cosh rz$$
$$- rz(A\cosh rz + B\sinh rz)\} \quad \text{.........(37)},$$

$$p_{xy} = -(\lambda+\mu) pq\frac{z}{r}(A\sinh rz + B\cosh rz)$$
$$+ \frac{pq}{r^2}[\{(\lambda+\mu) rh\coth rh - \mu\} A\cosh rz + \{(\lambda+\mu) rh\tanh rh - \mu\} B\sinh rz]$$
$$\text{.........(38)},$$

with symmetrical expressions for p_{yy} and p_{yz}.

It is easy to verify that these expressions satisfy the original equations of equilibrium
$$\frac{\partial p_{xx}}{\partial x} + \frac{\partial p_{xy}}{\partial y} + \frac{\partial p_{xz}}{\partial z} = 0 \qquad \text{.........(39)},$$
with two similar equations.

8·3. In the problem of the support of the earth's crust, the values of P and Q must satisfy two conditions. Suppose the pressure to be reduced by S over the surface $z = -h$. This surface will be depressed by a certain amount, and the reduction in the weight of the matter above this level must be allowed for in finding the stress after deformation. If then ρ be the density of the superficial matter, we must have
$$P + Q = S + g\rho\,(w)_{z=-h} \qquad \text{.........(1)}.$$
Similarly, the depression of the lower boundary by the amount $w_{z=h}$ exposes it to the additional pressure due to a height $w_{z=h}$ of the underlying fluid, whose density will be supposed to be ρ'. Hence the additional stress p_{zz} across the boundary will be increased by $-g\rho' w_{z=h}$. The negative sign is required because the stresses are tensions, not pressures. But this is the change in the stress on the matter constituting the boundary; after the displacement the vertical tension across the plane $z = h$ has to support the weight of the column of the crust below it, and therefore the increase of tension there is $g\rho w_{z=h}$ more than that on the matter at the boundary. Hence
$$P - Q = -g\,(\rho' - \rho)\, w_{z=h} \qquad \text{.........(2)}.$$

Now we have
$$\begin{aligned} w &= -\frac{\lambda+\mu}{\mu}\frac{\partial \chi}{\partial z} + w_1 \\ &= -\frac{\lambda+\mu}{2\mu} z\,(A\cosh rz + B\sinh rz) \\ &\quad + \frac{\lambda+2\mu}{2\mu r}(A\sinh rz + B\cosh rz) \\ &\quad + \frac{\lambda+\mu}{2\mu}\,.\,h\,(A\coth rh\sinh rz + B\tanh rh\cosh rz) \end{aligned} \qquad \text{......(3)}.$$

Hence
$$w_{z=-h} = \frac{\lambda+2\mu}{2\mu r}(-A\sinh rh + B\cosh rh) \qquad \text{.........(4)},$$
$$w_{z=h} = \frac{\lambda+2\mu}{2\mu r}(A\sinh rh + B\cosh rh) \qquad \text{.........(5)}.$$

On substituting these expressions in (1) and (2), and expressing P and Q in terms of A and B by means of 8·21 (33) and (34), we have
$$(\lambda+\mu)(A\cosh rh + Arh\,\mathrm{cosech}\,rh - B\sinh rh + Brh\,\mathrm{sech}\,rh)$$
$$= S + g\rho\frac{\lambda+2\mu}{2\mu r}(-A\sinh rh + B\cosh rh) \qquad \text{.........(6)},$$
$$(\lambda+\mu)(A\cosh rh + Arh\,\mathrm{cosech}\,rh + B\sinh rh - Brh\,\mathrm{sech}\,rh)$$
$$= -g\,(\rho' - \rho)\frac{\lambda+2\mu}{2\mu r}(A\sinh rh + B\cosh rh) \qquad \text{.........(7)},$$

which suffice to determine A and B. If we write

$$A - B = 2M; \quad A + B = 2N \qquad \text{.........(8)},$$

(6) and (7) take the alternative forms

$$(\lambda + \mu)\left\{Me^{rh} + Ne^{-rh} + \frac{rh}{\cosh rh \sinh rh}(Ne^{rh} + Me^{-rh})\right\}$$
$$= S + \frac{g\rho(\lambda + 2\mu)}{2\mu r}(-Me^{rh} + Ne^{-rh}) \qquad \text{.........(9)},$$

$$(\lambda + \mu)\left\{Ne^{rh} + Me^{-rh} + \frac{rh}{\cosh rh \sinh rh}(Me^{rh} + Ne^{-rh})\right\}$$
$$= -g(\rho' - \rho)\frac{\lambda + 2\mu}{2\mu r}(Ne^{rh} - Me^{-rh}) \qquad \text{.........(10)}.$$

Let us first consider the case where rh is large. Then (10) shows that Ne^{rh} and $4rhMe^{-rh}$ are of the same order; for we already know that $(g\rho)/(\lambda + 2\mu)\kappa$ is small in the problems we are considering, and *a fortiori* $\dfrac{g(\rho' - \rho)(\lambda + 2\mu)}{2(\lambda + \mu)\mu r}$ is small. Hence to a first approximation

$$N = -4rhMe^{-2rh} \qquad \text{.........(11)},$$

and substituting in (9) we have

$$M = Se^{-rh}/(\lambda + \mu) \qquad \text{.........(12)}.$$

Hence by 8·2 (7) $\qquad \delta = \dfrac{S}{\lambda + \mu}\{e^{-r(z+h)} - 4rhe^{-3rh+rz}\} \qquad \text{.........(13)},$

or if we take as a new coordinate the depth below the undisturbed surface, putting

$$z + h = z_1 \qquad \text{.........(14)},$$

and neglect the terms in e^{rz}, on account of their small coefficients, we find from 8·22 (35) to (38)

$$p_{zz} = (\lambda + \mu)rA(h + z)e^{-rz} + (\lambda + \mu)Ae^{-rz} \qquad \text{.........(15)},$$

$$p_{xx} = \left(\lambda - \frac{\mu p^2}{r^2}\right)Ae^{-rz} + \frac{(\lambda + \mu)p^2}{r}A(h + z)e^{-rz} \qquad \text{...(16)},$$

$$p_{zx} = -(\lambda + \mu)pA(h + z)e^{-rz} \qquad \text{.........(17)},$$

$$p_{xy} = (\lambda + \mu)\frac{pq}{r}A(h + z)e^{-rz} - \mu\frac{pq}{r^2}Ae^{-rz} \qquad \text{.........(18)}.$$

These can be shown by differentiation to satisfy the conditions of equilibrium. We have from (8) and (12)

$$(\lambda + \mu)Ae^{rh} = S \qquad \text{.........(19)}.$$

Hence the stresses reduce to

$$p_{zz} = S(1 + rz_1)e^{-rz_1} \qquad \text{.........(20)},$$

$$p_{xx} = S\frac{\lambda - \mu p^2/r^2}{\lambda + \mu}e^{-rz_1} + \frac{p^2}{r}z_1e^{-rz_1} \qquad \text{.........(21)},$$

$$p_{zx} = -Spz_1e^{-rz_1} \qquad \text{.........(22)},$$

$$p_{xy} = S\frac{pq}{r^2}\left(rz_1 - \frac{\mu}{\lambda + \mu}\right)e^{-rz_1} \qquad \text{.........(23)}.$$

8·4. If now S is a simple harmonic function of κx and $\kappa' y$, we may fix the origin so that
$$S = \nu \cos \kappa x \cos \kappa' y \qquad \text{.........(1)},$$
where ν is a constant. Then
$$r^2 = \kappa^2 + \kappa'^2 \qquad \text{.........(2)},$$
and
$$p_{zz} = \nu (1 + r z_1) e^{-r z_1} \cos \kappa x \cos \kappa' y \qquad \text{.........(3)},$$
$$p_{xx} = \nu \left\{1 - \frac{\mu \kappa'^2}{(\lambda + \mu) r^2} - \frac{\kappa^2 z_1}{r}\right\} e^{-r z_1} \cos \kappa x \cos \kappa' y \qquad \text{......(4)},$$
$$p_{zx} = \nu \kappa z_1 e^{-r z_1} \sin \kappa x \cos \kappa' y \qquad \text{.........(5)},$$
$$p_{xy} = \frac{\nu \kappa \kappa'}{r^2} \left(r z_1 - \frac{\mu}{\lambda + \mu}\right) e^{-r z_1} \sin \kappa x \sin \kappa' y \qquad \text{.........(6)},$$
from which the expressions for p_{yy} and p_{yz} can be obtained by considerations of symmetry. Now the three principal stresses at any point are the roots of the cubic equation in ϖ,
$$\begin{vmatrix} p_{xx} - \varpi & p_{xy} & p_{xz} \\ p_{xy} & p_{yy} - \varpi & p_{yz} \\ p_{xz} & p_{yz} & p_{zz} - \varpi \end{vmatrix} = 0 \qquad \text{.........(7)}.$$

This determinant is an even function both of $\cos \kappa x$, $\sin \kappa x$, $\cos \kappa' y$, and $\sin \kappa' y$. The stationary values of each stress-component with respect to variations in x and y occur where $\cos \kappa x$ and $\cos \kappa' y$ are each equal to zero or ± 1. The same may be expected to be true of the principal stresses and of the differences between them, and this has been found to hold in all special cases so far examined; but a general proof has not yet been constructed, on account of the heaviness of the algebra involved. It will, however, be assumed that it is in general true that the greatest stress-differences occur at points where $\cos \kappa x$ and $\cos \kappa' y$ are each zero or ± 1, and the following discussion will therefore deal only with these points.

It will facilitate discussion if the surface of the earth is considered marked out into rectangles by the two sets of lines where
$$\cos \kappa x = 0 \text{ and } \cos \kappa' y = 0.$$
Then the points of the surface where $\cos \kappa x$ and $\cos \kappa' y$ are ± 1 will be the centres of these rectangles. If $\cos \kappa x$ and $\cos \kappa' y$ are both 1 or both -1, the point considered will be one of maximum elevation; if only one of them is -1, the point will be one of maximum depression. The points where
$$\cos \kappa x = \sin \kappa' y = 0,$$
and those where
$$\sin \kappa x = \cos \kappa' y = 0$$
are the midpoints of the sides of the rectangles. Those where
$$\sin \kappa x = \sin \kappa' y = 0$$
are the corners of the rectangles. We therefore proceed to discuss the stress-differences at points vertically below these various points identified on the surface.

8·41. Considering first the centres of the rectangles, we have

$$p_{yz} = p_{zx} = p_{xy} = 0 \quad \text{.........(1).}$$

Hence the three principal stresses at any depth are p_{xx}, p_{yy}, p_{zz}. We see at once that the greatest of these in absolute value is always p_{zz}, and the greatest stress-difference is therefore either

$$| p_{zz} - p_{xx} | = \left(\frac{\mu\kappa'^2}{(\lambda + \mu) r^2} + \frac{r^2 + \kappa^2}{r} z_1\right) \nu e^{-rz_1} \quad \text{.........(2),}$$

or

$$| p_{zz} - p_{yy} | = \left(\frac{\mu\kappa^2}{(\lambda + \mu) r^2} + \frac{r^2 + \kappa'^2}{r} z_1\right) \nu e^{-rz_1} \quad \text{.........(3).}$$

We see that the maximum value of $p_{zz} - p_{xx}$ in a vertical line is given by

$$\frac{\partial}{\partial z_1}\left\{\frac{\mu\kappa'^2}{(\lambda + \mu) r^2} + \frac{r^2 + \kappa^2}{r} z_1\right\} e^{-rz_1} = 0 \quad \text{.........(4),}$$

which makes

$$z_1 = \frac{2\kappa^2 + \kappa'^2\lambda/(\lambda + \mu)}{r\,(2\kappa^2 + \kappa'^2)} \quad \text{.........(5).}$$

Similarly the greatest value of $p_{zz} - p_{yy}$ occurs at a depth

$$\frac{2\kappa'^2 + \kappa^2\lambda/(\lambda + \mu)}{r\,(2\kappa'^2 + \kappa^2)} \quad \text{.........(6).}$$

So far the magnitudes of κ and κ' have not been considered. Without loss of generality, therefore, we may suppose the axes turned so that the axis of y is in the direction of the greater wave length. Thus κ will be greater than κ'. Then

$$| p_{zz} - p_{xx} | - | p_{zz} - p_{yy} | = (\kappa^2 - \kappa'^2)\left(-\frac{\mu}{(\lambda + \mu) r^2} + \frac{z_1}{r}\right) \nu e^{-rz_1} \text{...(7).}$$

Thus $| p_{zz} - p_{xx} |$ will be greater than $| p_{zz} - p_{yy} |$ if the depth exceeds $\frac{\mu}{(\lambda + \mu) r}$, but the latter will be the greater if the depth is less than this critical value. Now the greatest value of $| p_{zz} - p_{yy} |$ occurs at a depth $\frac{2\kappa'^2 + \kappa^2\lambda/(\lambda + \mu)}{r\,(2\kappa'^2 + \kappa^2)}$. The condition for this to be greater than $\frac{\mu}{(\lambda + \mu) r}$ is

$$(\lambda + \mu)\, 2\kappa'^2 + \kappa^2\lambda > \mu\,(2\kappa'^2 + \kappa^2) \quad \text{.........(8),}$$

which is true, since λ is positive and at least equal to μ. Hence the greatest value of $| p_{zz} - p_{yy} |$ occurs at a depth where it is less than the value of $| p_{zz} - p_{xx} |$ at the same depth, and *a fortiori* is less than the value of $| p_{zz} - p_{xx} |$ at the depth where the latter reaches its maximum. Hence the greatest stress-difference below the centre of a rectangle is $| p_{zz} - p_{xx} |$ at the depth $Z = \frac{2\kappa^2 + \kappa'^2\lambda/(\lambda + \mu)}{r\,(2\kappa^2 + \kappa'^2)}$, its value there being $\frac{2\kappa^2 + \kappa'^2}{r^2} \nu e^{-rZ}$.

8·42. At the corners of the rectangles we have

$$p_{zz} = p_{xx} = p_{yy} = p_{zx} = p_{yz} = 0 \quad \text{.........(1),}$$

$$p_{xy} = \pm \frac{\nu\kappa\kappa'}{r^2}\left(rz_1 - \frac{\mu}{\lambda + \mu}\right) e^{-rz_1} \quad \text{.........(2).}$$

Hence $$\varpi = 0 \text{ or } \pm \frac{\nu\kappa\kappa'}{r^2}\left(rz_1 - \frac{\mu}{\lambda+\mu}\right)e^{-rz_1} \qquad \text{.........(3)},$$

and the greatest stress-difference is

$$\frac{2\nu\kappa\kappa'}{r^2}\left|rz_1 - \frac{\mu}{\lambda+\mu}\right|e^{-rz_1} \qquad \text{.........(4)}.$$

8·43. To compare this with the greatest stress-difference at an equal depth at a point of maximum elevation or depression, we must consider two separate cases, according as z_1 is greater or less than $\frac{\mu}{(\lambda+\mu)r}$. If

$$z_1 > \frac{\mu}{(\lambda+\mu)r},$$

the stress-difference at a corner is less than $\frac{2\nu\kappa\kappa'}{r}z_1e^{-rz_1}$. The coefficient of $\nu z_1e^{-rz_1}$ in this is $\frac{2\kappa\kappa'}{r}$, which is less than $\frac{2\kappa^2}{r}$, which is less than $\frac{2\kappa^2+\kappa'^2}{r}$, which is the coefficient of $\nu z_1e^{-rz_1}$ in the expression for $|p_{zz} - p_{xx}|$ vertically below the centre of a rectangle. The other term in the latter expression is always positive, and therefore the stress-difference at a corner at any depth greater than $\frac{\mu}{(\lambda+\mu)r}$ is less than the stress-difference at an equal depth below a centre.

If the depth is less than $\frac{\mu}{(\lambda+\mu)r}$, the stress-difference below a corner is not greater than $\frac{2\nu\kappa\kappa'\mu}{r^2(\lambda+\mu)}$, and attains this value at the surface. Below a centre $|p_{zz} - p_{yy}|$ is greater than $|p_{zz} - p_{xx}|$; it is at least equal to $\frac{\mu\kappa^2}{r^2(\lambda+\mu)}$, attaining this value at the surface. Hence if κ is greater than $2\kappa'$, the stress-difference at the corner at any depth is less than that at the centre at the same depth; but if κ is less than $2\kappa'$, the stress-difference at the corner is the greatest at the surface.

Let us now compare the stress-difference at the surface at a corner, when κ is less than $2\kappa'$, with the greatest stress-difference at any depth below a centre. The maximum stress-difference below a centre is

$$\frac{2\kappa^2+\kappa'^2}{r^2}\nu e^{-rZ}.$$

But the expression for Z shows that rZ is less than unity, and therefore this stress-difference is greater than $\frac{2\kappa^2+\kappa'^2}{r^2e}\nu$. The condition for this to be greater than that at the surface at a corner is therefore that

$$2\kappa^2+\kappa'^2 > 2e\kappa\kappa'\frac{\mu}{\lambda+\mu}.$$

If λ and μ are equal, this reduces to

$$2\kappa^2 - e\kappa\kappa' + \kappa'^2 > 0,$$

which is true, since $e^2 < 8$. Hence if the Poisson ratio of the matter of the earth's crust has the standard value of $\frac{1}{4}$, which makes λ equal to μ, the greatest stress-difference below a centre will be greater than the greatest at a corner.

8·44. Considering now the midpoints of the sides, we have if

$$\cos \kappa x - \sin \kappa' y = 0,$$

$$p_{zz} = p_{xx} = p_{yy} = p_{yz} = p_{xy} = 0 \qquad \text{.........(1)},$$

$$p_{zx} = \pm\, \nu\kappa z_1 e^{-rz_1} \qquad \text{.........(2)}.$$

Hence

$$\varpi = 0 \quad \text{or} \quad \pm\, \nu\kappa z_1 e^{-rz_1} \qquad \text{.........(3)},$$

and the greatest stress-difference is $2\nu\kappa z_1 e^{-rz_1}$. A sufficient condition for this to be less than the value of $| p_{zz} - p_{xx} |$ at the same depth below a corner is that 2κ shall be less than $(r^2 + \kappa^2)\, r$. On substituting for r in terms of κ and κ' we find that this becomes

$$4\kappa^2 (\kappa^2 + \kappa'^2) < (2\kappa^2 + \kappa'^2)^2 \qquad \text{.........(4)}$$

which is true unless κ' is zero. Thus the stress-difference below the midpoint of a side of this set is always less than that at the same depth below the centre. Since κ' is less than κ, we see that the stress-difference at the midpoint of a side of the other set is still smaller.

8·45. Collecting these results, we see that if κ is greater than $2\kappa'$, so that the wave length in one direction is more than twice that in the other direction, the stress-difference at any point vertically below a place of maximum elevation or depression is greater than that at the same depth anywhere else. The maximum value is at a depth Z, equal to $\dfrac{2\kappa^2 + \kappa'^2\lambda/(\lambda + \mu)}{r\,(2\kappa^2 + \kappa'^2)}$, and its amount there is $\dfrac{2\kappa^2 + \kappa'^2}{r^2}\,\nu e^{-rZ}$, and greater than $\dfrac{2\kappa^2 + \kappa'^2}{r^2 e}\,\nu$. If, however, κ is between κ' and $2\kappa'$, there will be regions of finite depth, around the points where the lines of zero elevation cross, where the stress-difference exceeds that at the same depth at the points of maximum elevation or depression. Even in this case, however, it will not with normal substances exceed the greatest stress-difference vertically below a point of maximum elevation. Thus if the strength of the material is the same at all depths, permanent deformation will take place first by a sinking of the greatest elevations and a rise of the greatest depressions. Since the greatest stress-difference in these circumstances is $| p_{zz} - p_{xx} |$, the failure will take place by the greater stress p_{zz} overcoming the smaller p_{xx}, so that the matter below a region of maximum elevation will be forced out parallel to the axis of x; in other words, in the direction of the shorter wave length. Flow in the direction of the longer wave length will occur only if the surface stress is applied so quickly that the stress-difference $| p_{zz} - p_{yy} |$ reaches the amount necessary to produce

deformation before flow due in the direction of the shorter wave length has had time to reduce the stress-difference $| p_{zz} - p_{xx} |$ below the limit necessary to produce yield. Flow in regions below the slopes will become possible only when extra stress has been thrown on these regions by the failure of materials below the summits and hollows to support their share.

If κ is between κ' and $2\kappa'$, and the matter near the surface is weaker than that below, the stress-difference at the corners near the surface may become enough to produce yield before that below the centres has reached the greater value necessary to produce yield at the greater depth. The same may happen even in the case of uniform strength if μ is greater than λ; but in the actual earth these quantities are nearly equal. Now the stress-difference at the corners arises from the component p_{xy}, which represents a shear across each of the planes $x =$ constant, $y =$ constant, and parallel to it. The tendency will therefore be for matter near the places of greatest slope to flow downhill.

8·5. The present problem is a generalization of one discussed by Sir G. H. Darwin*, as the limiting case of the stresses due to a spherical harmonic deformation of high order. His results refer only to the case where κ is zero, corresponding to a series of parallel mountain chains of infinite length. In this case our results 8·4 (3) to (6) reduce to

$$p_{zz} = \nu (1 + \kappa z_1)\, e^{-\kappa z_1} \cos \kappa x \quad \text{.........(1)},$$
$$p_{xx} = \nu (1 - \kappa z_1)\, e^{-\kappa z_1} \cos \kappa x \quad \text{.........(2)},$$
$$p_{yy} = \nu \frac{\lambda}{\lambda + \mu} e^{-\kappa z_1} \cos \kappa x \quad \text{.........(3)},$$
$$p_{zx} = \nu \kappa z_1 e^{-\kappa z_1} \sin \kappa x \quad \text{.........(4)},$$
$$p_{zy} = p_{xy} = 0 \quad \text{.........(5)}.$$

The equation giving the principal stresses therefore becomes, if

$$\varpi = \sigma \nu e^{-\kappa z_1} \quad \text{.........(6)},$$

$$\begin{vmatrix} (1 - \kappa z_1) \cos \kappa x - \sigma & 0 & \kappa z_1 \sin \kappa x \\ 0 & \dfrac{\lambda}{\lambda + \mu} \cos \kappa x - \sigma & 0 \\ \kappa z_1 \sin \kappa x & 0 & (1 + \kappa z_1) \cos \kappa x - \sigma \end{vmatrix} = 0 \quad \text{...(7)},$$

the roots of which are $\dfrac{\lambda}{\lambda + \mu} \cos \kappa x$ or $\cos \kappa x \pm \kappa z_1$. The first of these roots evidently corresponds to the component p_{yy}, flow due to which is impossible, since it is the only component capable of producing motion parallel to the axis of y, and is itself independent of y, so that the resultant force parallel to the axis of y acting on any portion of the crust must be zero. The two principal stresses in a vertical plane perpendicular to the ridges differ by $2\nu\kappa z_1 e^{-\kappa z_1}$. This is independent of x, so that the stress-difference is the same at all points at the same depth, whatever their horizontal distances from the nearest ridge. This has been seen to be untrue in the more general case. The maximum stress-difference occurs

* *Scientific Papers*, 2, 481–84.

at a depth $1/\kappa$, and is equal to $2\nu/e$. There is no stress-difference at the surface, in which respect again we have a disagreement with the case where κ' is not zero.

8·6. In the above discussion rh has been supposed large. The stress-components introduced by the pressure applied over the surface have been found to decrease rapidly with depth, when the depth exceeds a definite finite value of order $1/r$. Hence the fluid upon which the crust has been supposed to rest has practically no additional stress to support, and it was therefore to be expected that the reaction $P - Q$ across the boundary between the solid and the fluid would be negligible, as has been found to be the case. So long as the condition that rh is great is satisfied, the solution for the regions where the stress is comparable with the pressure applied over the surface is the same as that for a solid of infinite depth.

If, however, rh is small, the stresses at all points within the solid layer would be expected to be comparable, and the fluid may have to support a large fraction of the pressure applied over the surface. In other words, the crust may undergo considerable flexure, and the extra hydrostatic pressure where it is depressed into the fluid may nearly balance the pressure on the top. Within the fluid there is no stress-difference, but in the surface layer there will be. In comparing the two cases it will be useful to evaluate the stress-differences in the crust, so as to find out whether they exceed those in a deep crust or not. Coming now to the case of a floating crust, of small depth in comparison with the wave length of the deformation applied, we may approximate to 8·3 (9) and (10) by neglecting powers of rh higher than the first. They reduce to

$$2(\lambda + \mu) A = S + g\rho \frac{\lambda + 2\mu}{2\mu r} (-Arh + B) \quad \text{.........(1)},$$

$$2(\lambda + \mu) A = -g(\rho' - \rho) \frac{\lambda + 2\mu}{2\mu r} (Arh + B) \quad \text{.........(2)},$$

provided that $B(rh)^3$ is small compared with A. We know already that in the class of problems we are considering $g\rho$ is small compared to $(\lambda + 2\mu) r$. Now λ and μ are nearly equal, and in these two equations the ratios of the terms in A on the right to those on the left are respectively $\frac{g\rho(\lambda + 2\mu)}{4\mu r(\lambda + \mu)} rh$ and $\frac{g(\rho' - \rho)(\lambda + 2\mu)}{4\mu r(\lambda + \mu)} rh$. In each of these rh is small, by hypothesis, and the other factor is of the order of $\frac{g\rho}{2\mu r}$, which is already known to be small. Hence the terms in A on the right are both products of two small factors, and therefore can be neglected. With this simplification the equations give

$$A = \frac{\rho' - \rho}{2\rho'(\lambda + \mu)} S \quad \text{.........(3)},$$

$$B = -\frac{2\mu r S}{g\rho'(\lambda + 2\mu)} \quad \text{.........(4)}.$$

The condition for the validity of the approximation is that

$$(rh)^3 < A/B \qquad \text{.........(5)},$$

while

$$\frac{A}{B} = \frac{g(\rho' - \rho)(\lambda + 2\mu)}{4\mu(\lambda + \mu) r} \qquad \text{.........(6)}.$$

Taking $g = 1000$, $\lambda = \mu = 5 \times 10^{11}$, and $\rho' - \rho = 1$, all in c.g.s. units, we have

$$\frac{A}{B} = \frac{7}{10^{10} r} \qquad \text{.........(7)}.$$

Thus our condition gives

$$r^4 h^3 < 7 \times 10^{-10} \qquad \text{.........(8)},$$

and if we put h equal to 50 km., or 10^7 cm., this gives

$$r < 5 \times 10^{-8}/\text{cm.} \qquad \text{.........(9)}.$$

If the distribution of pressure over the upper surface is simple harmonic, this shows that the wave length must be at least 1200 km. to make the present solution valid.

Substituting in the expressions for the stress-components, and remembering that z cannot exceed h within the crust, we have on rejecting powers of rh above the first, in accordance with the approximation already made,

$$p_{zz} = 2A(\lambda + \mu) \qquad \text{.........(10)},$$

$$p_{xx} = A\lambda\left(1 + \frac{p^2}{r^2}\right) + B\lambda\left(1 - \frac{p^2}{r^2}\right) rz \qquad \text{.........(11)},$$

$$p_{zx} = 0 \qquad \text{.........(12)},$$

$$p_{xy} = \frac{pq}{r^2}\{\lambda A - (\lambda + 2\mu) Brz\} \qquad \text{.........(13)}.$$

If in particular

$$S = \nu \cos \kappa x \qquad \text{.........(14)},$$

we have

$$p_{xy} = 0 \qquad \text{.........(15)},$$

and the greatest stress-difference is

$$|p_{zz} - p_{xx}| = |\{2B\lambda rz - 2A(\lambda + \mu)\}\nu \cos \kappa x|$$
$$= \left(\frac{4\mu\lambda\kappa^2 z}{g\rho'(\lambda + 2\mu)} + \frac{\rho' - \rho}{\rho'}\right)|\nu \cos \kappa x| \qquad \text{.........(16)}.$$

If z has its maximum value h, and the quantities involved have the values already assumed, together with

$$\rho = 2{\cdot}5;\ \rho' = 3{\cdot}5,$$

the first term within the bracket is approximately 3, while the second is 2/7. Thus the greatest stress-difference in a thin crust floating on a fluid will occur at the bottom surface, and will be about 3ν below the places of greatest elevation and depression.

In the thick crust the greatest stress-difference at any depth was found to be $2\nu/e$. Hence the stress-differences in a thin crust may considerably exceed those in a thick one.

8·7. In the foregoing investigation two extreme cases have been considered, the thickness of the crust having been supposed in the one to be great, and in the other to be small, in comparison with the wave length of the disturbing pressure applied over the surface. In each case the distribution of the stress-differences through the crust has been found. In the case of a thick crust the maximum stress-difference is below the greatest elevations and depressions, and at a depth about $1/2\pi$ of the distance between consecutive ridges. Its amount is about $2\nu/e$, where ν is the maximum pressure applied over the upper surface. When the whole thickness of the crust is less than $1/2\pi$ of the distance between consecutive ridges, the crust bends down as a whole until the extra upward pressure of the fluid on the regions bearing the extra load almost balances the weight of the load. When this is so, the depressed regions are compressed above and stretched below, the opposite holding for the elevated regions. The deformation produces stress-differences which may exceed several times the maximum that can occur in a thick crust of similar materials; the greatest is at the bottom of the crust.

CHAPTER IX

The Theory of Isostasy

> "*I* could show you hills, in comparison with which you'd call that a valley."
> "No, I shouldn't," said Alice, surprised into contradicting her at last: "a hill *can't* be a valley, you know. That would be nonsense...."
> The Red Queen shook her head. "You may call it nonsense if you like," she said, "but *I've* heard nonsense, in comparison with which that would be as sensible as a dictionary!"
> LEWIS CARROLL, *Through the Looking-Glass.*

9·1. *The Behaviour of Matter under Shearing Stress.* It was shown in 6·6 and 6·61 that the matter of the earth's crust at a depth between 300 km. and 400 km. has had time to cool down since solidification by something of the order of 200°; while the matter at a depth of 700 km. can have cooled by only a few degrees, at the most, from the melting point at the pressure that actually prevails at that depth. We should therefore expect that, while the rocks at the surface possess the properties of ordinary solids, a gradual transition in properties would be associated with increase of depth, and that the matter at a depth of about 700 km. would behave almost as a fluid. Unfortunately, however, this is much too simple a statement of the problem. Several quite different properties are commonly thought characteristic of fluids, but are by no means invariably associated. Thus the use of the term 'fluid' without some preliminary discussion of what is meant by it in the particular context would be certain to lead to confusion. Some account of the properties of matter in its various physical states is therefore necessary.

In studying the development of the earth, especially in relation to its surface features, we shall be largely concerned with phenomena of change of shape, both temporary and permanent. Hence the physical properties of its constituents that we chiefly need to know will be the relations between the changes of shape that they undergo and the stresses that produce these changes. Other properties of matter, so far as they concern us, will do so only in a subsidiary way. Any classification of substances, to be useful in geophysics, must therefore be based primarily on their behaviour under deforming stress. Such a classification will be outlined in what follows. It does not follow completely any classification known to me, for two reasons. First, no single account I have seen appears to deal adequately with all the properties that require description; and secondly, most accounts are methodologically unsatisfactory in that they use for purposes of definition properties incapable of being tested experimentally.

The dimensions of a body may be altered in two ways. First, it may be compressed or extended by a uniform pressure or tension over the whole of its surface. This property is called *compressibility*. In many substances,

which are called *isotropic*, the dimensions in all directions alter in the same ratio when the pressure is uniform and normal to the surface. If, however, the pressure is not uniform, or if the substance as tested by the above criterion is not isotropic, the dimensions in different directions will in general change in different ratios. Any alteration of size or form in a body can be represented as a combination of changes of volume without change of shape, and changes of shape without change of volume. The simplest type of the latter occurs when a rectangular block has one face clamped, while a tension is applied in the plane of the opposite face. The block will be distorted, its angles ceasing to be right angles, but the volume will remain unaltered. A stress that alters angles without altering the volume is called a *shearing stress*. If, again, the tension over each face of the block is normal and uniform, and if the tensions over opposite faces are equal, but those over adjacent faces are different, the block will in general be altered both in volume and in shape, becoming most extended in the direction of the greatest tensions. The angles between lines inclined to the edges of the block will be altered, indicating the presence of shear. If three mutually perpendicular lines meeting in a point in the body remain perpendicular after the deformation, they are called *principal axes of the strain* at that point. In the case of the rectangular block just considered, lines through any point parallel to the edges are principal axes of the strain. Again, if the stress across a plane at any point is wholly normal to that plane, this plane is called a *principal plane of stress* at that point. At any point there are three mutually perpendicular principal planes of stress, and the tractions across them are called the *principal stresses* at the point. If the substance is isotropic, the principal stresses act along the principal axes of the strain. The difference between the greatest and least of the principal stresses is called the *stress-difference.* It is evident that in an isotropic substance the vanishing of the stress-difference indicates that there is no distortion at the point, the expansion or contraction being the same in all directions. Thus stress-difference and shear in an isotropic substance are always associated.

9·11. *Permanent Set.* Suppose now that a body is deformed in any way, and that the external deforming stresses, after being applied for some time, are removed. Further changes of form will in general follow. The rate of variation of shape may diminish with time in such a way that it is legitimate to infer that it is tending to zero and that the extent of the change of form from the original state is tending to a definite limit. If, for instance, the changes in several successive equal intervals of time decrease like the terms of a decreasing geometrical progression, such an inference will be justified. If the limit is different from the original state, the substance is said to have undergone *permanent set.*

Permanent set follows very different laws in different substances.

There may be a limiting stress-difference such that no permanent set is observed unless the stress-difference actually applied exceeds this limit, but such that any greater stress-difference always produces permanent set. This is usually called the *limit of perfect elasticity*, but the term is a bad one, since a body may be strained so as to show no permanent set, while being very far from perfectly elastic in another sense, which will be explained below. It will here be called the *set-point*. It may happen that the set-point is not appreciably different from zero; this is true not only for typical fluids, but for such metals as cast iron. Often, however, it is different from zero.

The nature of the set varies in different cases. It may involve a continuous deformation such as a liquid undergoes, particles originally in contact remaining in contact; or the constituent parts may roll or slip over one another, with or without internal change of form; or parts originally within the body may cease to be in contact with other parts of the body at all. The first type may be called *flow*, the second *elastic hysteresis*, and the last *fracture*; but the property of greatest geophysical interest is the same in all, namely that the substance acquires a permanent elongation in the direction of greatest tension (or least pressure) and a permanent contraction in that of least tension (or greatest pressure); and we are fundamentally concerned with the existence and amount of the set, and not with its nature or the details of the process that brings it about.

The amount of the set may depend not only on the stress-differences applied, but also on how long the stress-difference has exceeded the set-point. It may happen that the set is practically independent of this time; this property is connected with elastic hysteresis. In this case the shape of the body tends to a definite limit under the stress applied, and does not surpass the limit, however long the application continues, unless the stress is further increased.

9·12. *Plasticity and Strength.* When the stress is sufficiently great, however, it will be found that the rate of change of shape shows no sign of falling off when the stress is applied for a long interval. If this is so, the extent of the recovery when the stress is removed is practically independent of the time of application of the stress, so that the set is an increasing function of the time of application, and could theoretically be made to surpass any limit by increasing the time sufficiently. The last property is here called *plasticity*. It is one of the most important properties of matter, since the flow of fluids, the malleability and ductility of some solids, and the brittleness of other solids, are all particular cases of it. The critical stress-difference, above which the rate of change of shape does not decrease when the time of application of the stress increases, may be called the *strength* of the material; and one substance may be

said to be *weaker* than another if it has a smaller strength. Every substance can be made to show plasticity by exposing it to a sufficiently great stress-difference; a body heterogeneous in constitution may, however, show it only when the stress-difference has become great enough to produce it in every constituent*.

The ratio of the stress-difference to twice the rate of shear during plastic flow may be called the *viscosity*. If a body has been undergoing deformation through plasticity, and the external stresses are gradually diminished, the plastic deformation in any element will cease when the stress-difference there sinks to the strength, and thenceforward no further set will be acquired until the stress-difference again surpasses the strength of the material.

9·13. *Rigidity and Elastic Afterworking*. If a body is exposed to a stress-difference insufficient to produce permanent set, and is then released, it may oscillate about its original position, the extent of the oscillations gradually diminishing to zero, or it may return to its original position as a limit without ever passing through it. In either case the tendency to return is said to show *rigidity*. Rigidity and strength are quite distinct properties, but are habitually confused in geological literature. A substance that oscillated about the original position, the oscillations retaining permanently the same mechanical energy, would be called *perfectly elastic*; but such a substance does not exist. The concept of a perfectly elastic body is, however, useful, for in certain circumstances the behaviour of real bodies approximates very closely to that of perfectly elastic ones, which therefore serves as a valuable standard of comparison, and is, at the same time, susceptible of exact mathematical treatment. The dying down of the deformation is called *elastic afterworking*. An example of this phenomenon is afforded by certain biscuits containing treacle, notably ginger snaps, when slightly stale. If bent to an extent insufficient to start a crack, and then released, the biscuit may be seen to creep back slowly to its original flat form.

9·14. If a substance that shows rigidity and elastic afterworking when its stress has not exceeded the set-point is afterwards exposed to stress-differences greater than the set-point, and is afterwards released, its behaviour during the stress and recovery is a complex process, for rigidity,

* If plastic deformation commences in one constituent, the failure of this constituent to bear the stress-difference falling on it will throw additional stress-difference on the others. An example of this was seen in the problem discussed in 8·6. Hence a plastic yield of the body as a whole will take place more readily than it would if the body were composed entirely of its strongest constituent. This fixes an upper limit to the strength of a heterogeneous body as a whole. The possibility of thus fixing an upper limit requires some emphasis, since it is sometimes thought that the heterogeneity of the earth imposes an insuperable obstacle to the application of any elastic theory. Even if constituents in a state of mechanical mixture become separated, this statement will remain true.

elastic afterworking and set all take part. It will therefore return only part of the way, if at all, towards its original configuration, and the return will be slow, the oscillations, if any, gradually dying down. To disentangle these effects, and to say how much of the motion is due to each, would require more experimental investigation than has yet been carried out.

9·15. *Definitions of Solids and Fluids.* Typical fluids have no strength and no rigidity. Since elastic afterworking arises only as a property of rigidity, this property also is absent from fluids. Suppose, for instance, that a horizontal plate is suspended in a liquid, and is then moved horizontally through it. The surface remains level, so that gravity does not affect the motion. It is found that, however small the stress acting on the plate may be, it always deforms the fluid, particles originally in vertical columns acquiring horizontal displacements with regard to one another; and that when the plate is no longer acted upon by stress the fluid merely comes to rest, showing no tendency to return. These properties have been defined to be the criteria for absence of strength and rigidity, which are therefore zero in fluids.

But although typical fluids possess neither strength nor rigidity, these properties are not invariably associated. Shoemaker's wax, for instance, is a famous example to the contrary. It is possible to make tuning forks of it, whose free vibrations have a frequency sufficiently high to enable them to give out an audible note; the resilience thus indicated implies rigidity. Yet when one of these forks is left to itself, it gradually flows out under its own weight, until a uniform flat surface has been produced. Hence it has no strength*. In general solids possess both strength and rigidity, but both properties diminish rapidly as the temperature approaches the melting point, and disappear as the substance melts. Strength is usually the first to show great diminution, and in the case of many glasses has quite disappeared some hundreds of degrees below the melting point. Thus to use the term 'fluid,' without carefully specifying whether absence of strength or absence of rigidity is the defining characteristic, would be certain to lead to errors. Each convention would have its advantages, and both have been used for the purpose. Absence of strength was used as the defining quality by Lord Kelvin, who has been followed by most physical writers. The definition, however, does not appear to have been followed in practice. If we are to have a coherent scheme the definition that should be used should be that actually used in melting point determinations, which has quite another basis. In such determinations the substance is not kept at a uniform temperature for as long a time as would be required, for instance, to make pitch at ordinary temperatures acquire a flat surface. Thus whatever the practical criterion of the fluid state may be, it is not absence of strength. Strength, indeed, is

* Lord Kelvin, *Baltimore Lectures*, Camb. Univ. Press, 1904, 9–10.

lost by all impure substances, and by many pure ones, at temperatures far below the accepted melting points. The property actually used is the acquirement of mobility; that is, the substance is considered to be fluid when it can be poured. Now if one tries to pour a substance possessing rigidity, the flow is resisted by rigidity until the stress-difference surpasses the strength, and afterwards by the resistance to plastic flow, so long as this continues; thus mobility implies the smallness of both these qualities. The absence of rigidity in the liquid state is again shown by the absence of any tendency towards elastic recovery of form when the substance has been poured. It therefore appears that the characteristic property of fluids is twofold: they have zero rigidities, and their viscosities are small in comparison with those of the same substances in the solid state. Either of these properties might be used as an expression of the practical criterion, the great reduction in viscosity on melting being probably the more convenient; but they are closely associated in actual substances, and it will be a matter for no surprise if rigidity in a substance is found to be present almost or quite up to the melting point. We shall therefore recognise the fluid state of a substance by the absence of rigidity and by the smallness of the viscosity in comparison with that in other states of the same material. Other states characterised by high viscosity, with or without rigidity and strength, will be called solid. Thus pitch at ordinary temperatures will be regarded as a solid.

9·16. *Properties of Solids.* It will be seen that this definition of the solid state suggests a further classification of solids according to their possession of rigidity and strength. It is possible, though not certain, that all solids possess rigidity. They may, however, be devoid of strength.

There is only one state of solids in which they are quite lacking in strength. In this state they are amorphous and practically uniform throughout. There is another state, however, bearing, at first sight, a close resemblance to the last, in so far as solids in this state are also amorphous and uniform, but differing from it in the possession of considerable strength. Both states are commonly described as *vitreous* or *glassy*. The difference between them is so important, however, as to merit a difference in nomenclature. The former will here be called the *liquevitreous* and the latter the *durovitreous* state. Any given vitreous substance is of the former type above a certain critical temperature, and of the second type below that temperature. This critical temperature is, of course, to be carefully distinguished from the critical temperature encountered in the discussion of the transition between the liquid and gaseous states. This is true for instance of all ordinary kinds of glass and of fused silica. The transition is usually gradual. Thus in one kind of hard glass with a critical temperature of 750° the strength below 750° has been found to vary roughly as $(750 - V)^2$, where V is the temperature. In a soft

glass the critical temperature may be something like 400°–450°. All vitreous substances are isotropic.

Another type of solid state is the state characterised by the existence of a definite crystalline form. The crystalline state is inherently weaker than the durovitreous, because every crystal possesses cleavage planes, over which slip may take place for stress-differences much less than are required to produce plastic flow in the durovitreous state. Dr A. A. Griffith informs me, for instance, that he has maintained vitreous silica at room temperature for a week at an elastic extension of 5 per cent., without detecting any flow. If there had been a flow of 0·005 per cent. of the length of the fibre he could have detected it. This corresponds to a far greater stress-difference than could be withstood for so long by any substance in the crystalline state.

Only crystals of a small class are isotropic, namely those belonging to the cubic system; and even in these the isotropy disappears when they have been strained so as to produce internal slip. In actual rocks, however, the crystals are orientated in all directions and usually are not even all of one kind. Hence the differences in elastic properties in different directions shown by different crystals will be expected to balance one another, and so long as we are dealing with masses of rock so large as to include a very large number of crystals, we may regard them as isotropic.

Vitreous substances often tend to crystallize when near particular temperatures, which may be above or below the critical temperature. If the temperature of crystallization is above the critical temperature, crystallization is resisted only by viscosity, and therefore must occur if sufficient time is allowed. If, however, it is below the critical temperature, the strength of the durovitreous matter may be enough to withstand the stresses involved in crystallization, and crystallization will then be impossible. The strength increases rapidly with decrease of temperature; hence if a substance is cooled sufficiently quickly through the liquevitreous state and the hotter part of the durovitreous state, crystallization may not have time to start, and may be afterwards permanently prevented.

The lack of strength of liquevitreous substances, combined with the use of the same name for both them and durovitreous substances, has led to a widespread belief that all substances of both classes are 'supercooled liquids' and able to flow to an indefinite extent under any stress-difference, however small, provided it is maintained for a sufficient time. The latter statement is true of liquevitreous substances, but not of durovitreous ones; the former depends on the definition of 'supercooling,' and will not arise in the present work. Ordinary window glass, with no horizontal pressure to support it, resists the stress due to its own weight for hundreds of years without appreciable set, a longer time than has been available in any experiments supporting the statement that all glasses are devoid of strength.

In addition to the above types of solid, there are two types of fragmentary solid. Some solids are composed of a vast number of small particles, which may be fastened together by a matrix or quite free. In the former case the solid follows the rules for mixtures already elaborated. The only need for special warning is that the matrix may be stronger than the particles; as may be seen from an inspection of an ordinary piece of broken brick concrete, where the fracture goes through the fragments of brick instead of through the cement or along the boundaries between brick and cement. If there is no matrix, strain again follows the usual rules until it becomes great enough to make the particles roll or slide over one another, after which the substance behaves more or less like a solid showing elastic hysteresis and plasticity together.

9·2. *Probable Mechanical Properties of the Earth's Crust.* Coming now to the application to the earth of this account of the properties of matter, we notice that rigidity, viscosity, and strength all increase as the temperature falls from the melting point, but that whereas strength may remain zero until cooling through some hundreds of degrees has taken place, viscosity is certainly, and rigidity probably, considerable as soon as the substance is below the melting point. If the substance crystallizes, the strength will become appreciable if the crystallization is complete and if the substance was previously in the liquevitreous state; but if it has already attained the durovitreous state, crystallization will weaken it to some extent. The cooling within the earth is so slow that it must ensure crystallization if the crystallizing temperature is reached. In any case, however, we shall expect that the rocks of the earth's crust, where they have cooled by only a few degrees since solidification, will have no strength, but that where they have cooled by some hundreds of degrees their strength will be considerable. The question is complicated by the high pressures prevailing at considerable depths in the earth's crust; but it will be observed that by hypothesis the initial temperature at any depth was the melting point appropriate to the pressure at that depth, and that just below the melting point every substance is lacking in strength, but has by definition considerable viscosity and probably considerable rigidity. These statements are true whatever the pressure; all we need to assume is that the temperature interval between solidification and the acquirement of strength remains of the same order of magnitude at the pressures existing at depths of some hundreds of kilometres as it is at the surface. If this plausible hypothesis be granted, it follows that the rocks at a depth of 700 km. can have no appreciable strength, that those at a depth of 300 km. may be just acquiring it, and that those at a depth of 100 km. are probably very strong. Rocks at all depths, however, may have great rigidity. The rigidity may be greater than that of rocks of the same constitution at the surface; for the rigidity is raised by pressure,

and this effect may be enough to counteract the reduction due to the greater temperature.

9·21. Let us consider the stresses in the crust due to the weight of a series of parallel ranges of mountains. We saw in 8·5 that if the elastic constants (i.e. the incompressibility and the rigidity) are the same throughout the crust, and if the depth of the crust is more than, say, half the distance between consecutive ridges, the maximum stress-difference will occur at a depth approximately equal to $1/2\pi$ of the distance between consecutive ridges, and will be equal to $2/e$ times the weight of the load per unit area where it is greatest. Now consider a series of ranges 3 km. in height. Taking the density as 2·7, we find the maximum stress-difference to be 6×10^8 dynes/cm.^{2} The crushing strength of basalt, which is probably typical of the rocks at depths of 30 km., is $1{\cdot}2 \times 10^9$ dynes/cm.^{2} Thus if the earth's crust were uniform and of the same strength as basalt in the laboratory, the weight of mountain ranges comparable with the Rockies and the Alps would be insufficient to produce plastic deformation of the crust at any depth; the weight of the Himalayas might be just sufficient.

9·22. When the reduction of strength with depth is taken into account, this statement evidently requires some qualification. If we assume as a working hypothesis that the earth's crust at depths greater than 400 km. possesses no strength, it will follow that any inequality capable of producing appreciable stress-differences at this depth in a uniform crust must produce plastic deformation in the actual crust. The weak matter will flow out laterally, and the upper crust will bend down, the amount of the depression being determined by the elasticity of the crust and the density of the weak matter, in accordance with the discussion of Chapter VIII. In the process the stress-differences in the upper crust will be increased, becoming greater than the maximum stress-difference in a uniform crust. If the distance between consecutive ridges is large in comparison with the depth of the layer of weakness, the approximation of 8·6 will become applicable. In this case the extent of the depression is such that the weight per unit area of the weak matter that flows out is nearly equal to the pressure applied over the surface, which is itself equal to the weight of added matter per unit area. Thus the effect of flow in a thin crust will be that the total quantity of matter in a column of given cross section down to a place in the layer of weakness at a given depth below the original surface will be unaltered.

9·23. The discussion of the last paragraph is still too simple an account of the facts. In the actual crust we should not expect that the rocks of the crust would have a uniform strength down to a certain depth, and no strength below that depth. Thus the last result does not represent the

true state of affairs, but it provides a basis for further investigation. We should expect a gradual variation of strength, possibly distributed over some hundreds of kilometres in depth; for the extent of cooling since solidification decreases continuously with depth, while the pressure increases continuously. Even in glasses of uniform composition the reduction of strength in passing from the durovitreous state to the liquevitreous state is gradual, and in the earth's crust the change must be gradual both for this reason and because rocks are in general mixtures. Hence a continuous decrease of strength would be expected, starting at the layer of greatest strength, whose depth is so far unknown, and finishing at a depth of about 400 km., below which strength is probably absent. The consequence will be that additional pressure applied over the surface will produce plastic deformation, not only in the region of zero strength, but also in all places above it where the stress-differences produced exceed the strength. Thus it is possible that the major part of the flow involved in the adjustment of the earth's crust to superficial inequalities may take place at depths much less than 400 km. If so, the crust may be treated as thin when the horizontal extent of the inequalities considered is much less than the 3300 km. given by 8·6 (8) for a crust of this thickness.

9·24. It will be noticed that the uniformity of mass per unit area over the earth's surface, inferred to hold when the crust is thin, would imply the absence of any superficial inequality if the density of the added matter was equal to that of the matter where flow first occurs; for to produce this uniformity the depth of matter that flows out would have to be equal to that of the added matter, and thus the depth of the matter in a vertical column would be the same as before. Hence no superficial inequality could persist if the crust were thin and the density uniform. If, however, the lightest matter was on top, when the adjustment was complete the depth of matter that flowed out would be less than that of the added matter, and then a projection would remain on the surface; but the greater part of the added matter would sink below the original surface in the process, and the projecting portion would correspond only to the visible portion of an iceberg, the part below the surface corresponding to the much larger portion of the iceberg that lies below the surface of the water.

9·25. The discussion of Chapter VIII referred to a flat earth of infinite extent. The results are readily adapted to give an interpretation that takes account of the finite size of the earth. It was seen that the stress-differences decreased rapidly with depth, so that a depth could always be found such that the stress-differences below it produced by a given load on the surface did not exceed any assigned limit. Thus the hydrostatic state, in which the stress-differences are zero, is approximately undisturbed below a certain depth. Now in the actual earth, if the strength below a certain depth is zero, the hydrostatic state must exist below that depth and be

unaffected by surface load. An important property of the hydrostatic state is that the surfaces of equal density, equal gravitation potential, and equal pressure coincide. Our condition is therefore that the pressure, density, and gravitation potential below the layer of weakness are unaffected by surface load; and if the load is of such an extent that the crust can be regarded as thin, the mass per unit cross section in a column, cut down to some definite equipotential surface within the layer of weakness, will be unaffected by surface load, and will be the same for all places.

9·26. It has been seen that a great difference is to be expected between the adjustment of the earth's crust, on the one hand, to inequalities whose horizontal extent is so small that the stress-differences they produce reach their maxima far above the layer of weakness, and on the other hand to inequalities whose horizontal extent is large. In the former case the flow produced is inappreciable, and hardly affects the mass in a vertical column; any matter added, or any valley denuded away, will produce no flow reducing the effect on gravity outside the solid substance of the earth. In the latter case, the addition of extra mass to the surface will produce a disturbance to gravity in its neighbourhood, but the reduction of mass below will cause the anomaly to be much less than if the earth was undeformable. A great deal of observational evidence has been acquired concerning the effect of surface inequalities on gravity, and this will now enable quantitative tests to be applied to the theory so far developed.

9·3. *Effect of Uncompensated Surface Inequalities on Gravity.* The older geologists and geodesists regarded mountains as composed of matter of much the same density as the rest of the crust, and it was not realized that their weight would be expected to produce any deformation of the material below them, nor that the density of the matter below a mountain range might differ systematically from that of the matter at an equal depth below a plain or even an ocean. Now if a mountain is considered merely as an extra mass superposed on a previously uniform crust, and its deforming effect on the interior is ignored, it is possible to compute its gravitational attraction on bodies in its neighbourhood. The attraction can also be found experimentally, and the result compared with that calculated. The experiment was carried out on several mountains during the eighteenth century, but the results were in conflict.

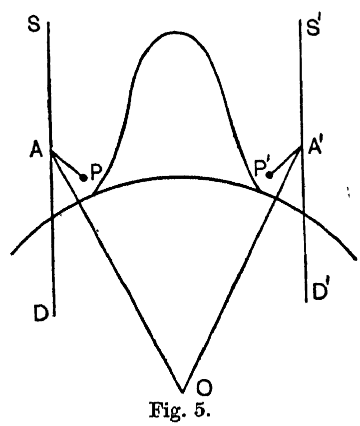


Fig. 5.

The principle of the measurement of the attraction of a mountain may be illustrated by means of the above figure. Let A and A' be two pivots on opposite sides of the mountain, from which two pendulums are sus-

pended. Let O be the centre of the earth, supposed spherical. Let AS, $A'S'$ be lines joining A and A' to a fixed star in the plane AOA', and let AD, $A'D'$ be their prolongations past A and A'. Then the angle AOA' is nearly AA'/R, where R is the radius of the earth. Now AA' can be found from surveying operations, and therefore AOA' is determinable. Let its value be α. But we have, since AS and $A'S'$ are parallel,

$$OAD + OA'D' = AOA' = \alpha.$$

Next, the zenith at A, say, is defined to be on the prolongation of the plumb-line past its point of support. Hence the zenith distance of the star is equal to PAD. But the zenith distance is observable. Thus we can find PAD and $P'A'D'$ from observation. Let their sum be β. Then we have by subtraction

$$OAP + OA'P' = \beta - \alpha.$$

But if the mountain had no deflecting effect, the two pendulums would point straight towards the centre of the earth, and therefore OAP and $OA'P'$ would both be zero. Hence $\beta - \alpha$ is equal to the sum of the deflexions of the plumb-line at the two stations owing to the attraction of the mountain.

Now if g be the intensity of gravity, and the bob of the pendulum AP be exposed to a small horizontal acceleration γ due to the mountain, the deflexion of the pendulum is, by the ordinary rules of statics, γ/g. Hence if γ' refer to the other pendulum,

$$\gamma + \gamma' = g\,(\beta - \alpha).$$

The accelerations γ and γ' are calculable from the law of gravitation, if the region has been surveyed, and therefore this relation affords a test of the theory.

The first attempt to use this method was apparently made by Bouguer*. The sum $\beta - \alpha$ of the deflexions due to the mountain Chimborazo was found to be very much less than that calculated from the law of gravitation. At that time the constant of gravitation was known only vaguely, but the deflexion of the plumb-line was much less than could be reconciled with any reasonable value. Maskelyne, in 1774, repeated the experiment at Schiehallion, in Perthshire. His results gave

$$\alpha = 41''; \quad \beta = 53''.$$

The value of $\gamma + \gamma'$, namely

$$\gamma + \gamma' = 0{\cdot}06 \text{ cm./sec.}^2$$

inferred from this, was used to estimate the constant of gravitation, and hence the mean density of the earth. The density was found to be 4·71. This is decidedly lower than modern estimates, but the discrepancy is much less than that found by Bouguer. A repetition by Petit† in the Pyrenees showed that their attraction was not only small, but actually negative; the plumb-line appeared to be deflected away from the moun-

* *La Figure de la Terre*, Paris, 1749. † *Comptes Rendus*, **29**, 1849, 729–34.

tains. Indeed the attraction of mountains was generally found to be nearer to zero than to the values calculated on the supposition that the underlying matter was of normal density. Schiehallion was an exception to the general rule; the reason it was exceptional was probably that it was of smaller size and therefore unable to produce much deformation below.

9·31. Another method of testing the attraction of a mountain is to consider the intensity of gravity on top of it, instead of the direction at its sides. For this purpose it will be sufficient to regard the mountain as flat; this is justifiable, for the width of a mountain or range of mountains is always much greater than its height. If the mass of the earth be M, its mean density ρ_0 and its radius a, the intensity of gravity at the surface is

$$g = fM/a^2 = \tfrac{4}{3}\pi f\rho_0 a \qquad \text{.........(1)},$$

where f is the constant of gravitation. At a height h above the surface, in the free air, the intensity of gravity is

$$\frac{fM}{(a+h)^2} = \frac{fM}{a^2}\left(1 - \frac{2h}{a}\right) = g\left(1 - \frac{2h}{a}\right) \qquad \text{.........(2)},$$

h^2 being neglected. If instead of the interval between sea-level and height h being filled with air, it is occupied by a mountain of density ρ', it is known from the theory of attractions that it will add an amount $2\pi f\rho' h$ to the attraction above it. Hence the total intensity of gravity at the top of the mountain, on the hypothesis again that the matter below it is of normal density, is

$$g\left(1 - \frac{2h}{a}\right) + 2\pi f\rho' h = g\left(1 - \frac{2h}{a} + \frac{2\pi f\rho' h}{g}\right)$$
$$= g\left(1 - \frac{2h}{a} + \frac{3}{2}\frac{\rho' h}{\rho_0 a}\right) \qquad \text{.........(3)}$$

by (1). Hence the excess of gravity at the top of a mountain over its value at sea-level is $-\frac{2gh}{a}\left(1 - \frac{3}{4}\frac{\rho'}{\rho_0}\right)$. In actual cases ρ' is always less than ρ_0, and therefore this anomaly is always negative. In other words, the gravity on a mountain top is less than elsewhere. This formula was obtained by Bouguer and by Young. As in the case of the deflexion of the pendulum, it was found when tested to give considerable errors; the actual anomaly of gravity on the top of a mountain was nearer to $-2gh/a$ than to the Bouguer formula.

9·4. *Compensation*. It is seen from an examination of the formula for the gravity anomaly that only the second term arises from the attraction of the mountain itself. Thus the statement that the gravity anomaly on the top of a mountain is equal to $-2gh/a$, implies that it is the same as would be correct if the mass of the mountain were zero, its height remaining the same. The hypothesis that the mountain is an egg-shell

gives better accordance with truth than the hypothesis that its weight does not deform the interior. A hypothesis virtually identical with the former was offered by Boscovich* when he suggested that mountains were swellings caused by the earth's internal heat, no extra matter being added. The use of this formula was reintroduced by Helmert, by whose name it is generally known.

Thus with regard both to the deflexion of the plumb-line and to the intensity of gravity, the attraction of a mountain is very much less than it would be if a deformation were not produced below. But this is exactly what would be expected from the theory of the deformation of the layer of weakness by superficial load. It has been seen that the mass under a given area lost in any deformation of a thin crust lying on a weak interior is nearly equal to that added on top. The two effects balance, so that the mass in a vertical column of unit cross section is unaltered, instead of being increased by the mass per unit area of the mountain added, or decreased by the mass per unit area denuded away from a valley. Hence the extra attraction in any direction due to a mountain is partly neutralized by the diminution in the attraction due to the loss of mass below it. The failure of the Bouguer formula to fit the facts, and the rough agreement with fact of the Helmert formula, are therefore in accordance with the theory so far elaborated of the strength of the earth's interior.

9·41. The hypothesis that the inequalities of mass over the surface are neutralized by opposite inequalities at an appreciable depth, as distinct from the Boscovich-Helmert hypothesis that mountains are of zero mass, is due to Archdeacon Pratt†, who reached it from the gravity observations, and not by way of cosmogony, as has been done in this work. The name 'isostasy' was coined for this hypothesis by Major C. E. Dutton in 1889. The term 'compensation' is also used. Some writers use 'compensation' to denote the fact of approximate uniformity of mass over the earth within a vertical column of constant cross section extending down to a standard equipotential surface, and restrict 'isostasy' to the physical process that leads to the establishment of this state.

It will be seen that our theory concerning the strength of the earth's crust predicts the existence of compensation as a first approximation to the truth; it does not assert that compensation must in all cases be exact. The existence of compensation has been inferred on the supposition that the inequalities considered have a horizontal extent large in comparison with the depth of the layer of zero strength, and probably over 2000 km. We therefore expect inequalities of this extent, however small in height, to be completely compensated.

* *De Litteraria Expeditione per Pontificiam Ditionem*, 1750, p. 475; or Todhunter, *Mathematical Theories of Attraction*, **1**, 313.

† *Phil. Trans.* **149**, 1859, 779–796; **161**, 1871, 335–357.

9·42. Inequalities of smaller horizontal extent, on the other hand, will produce a depression over the layer of zero strength too small to give complete compensation; the depression for such inequalities will follow a law intermediate between that correct for widespread inequalities and that correct for inequalities whose extent is small compared with the depth of the layer of zero strength, and we know that in the latter case there is no compensation. When adjustment in the layer of zero strength has taken place, however, some stress-difference will continue to exist in the weak layers above, and may be enough to cause plastic deformation in them. This must be the case if the strength decreases continuously downwards. For some stress-difference, different from zero, must exist in the crust both before and after yield has occurred in the layer of zero strength; and if the strength of the crust tends continuously to zero as we approach the layer of zero strength, there must be a region above this layer where the stress-difference exceeds the strength, and will therefore produce plastic deformation. Thus plastic deformation will gradually spread upwards. Since, however, adjustment is now taking place in a region of finite strength, plastic deformation will cease when the stress-difference at the depth considered has sunk to the strength of the material, which will necessarily be before it has become zero. Hence the depression will be less than it would be if the crust below the layers bent without plastic deformation were in a state of hydrostatic pressure. Thus we should expect that the compensation of inequalities of horizontal extent small compared with 2000 km. would be incomplete. The extent of the imperfection of compensation should depend only on the stresses in the underlying matter, and therefore cannot exceed a limit depending on the strength of this matter. It therefore renders possible the determination of a lower limit to the strength of the earth's crust over a considerable range of depth.

9·5. *Hayford's Hypothesis*. The most detailed work on the compensation of surface inequalities so far achieved is that of the United States Coast and Geodetic Survey*. The reductions of the observations have been made independently of any physical theory regarding the cause of the compensation, but, in order to have a definite hypothesis capable of quantitative test, Dr Hayford, when head of the Survey, introduced a special hypothesis about the nature of the compensation. The hypothesis is that under an elevated region the density at all depths falls short of the average for the same depth by the same amount, provided the depth does not exceed a certain maximum, and that at greater depths than this the density is the same over any equipotential surface. A physical interpretation of this hypothesis is easily obtained. Suppose that the density of the matter composing the elevation is ρ', and that the mean elevation of its surface above the mean surface of the land would, in the absence of

* The Indian Survey, under Sir S. G. Burrard and Sir G. P. Lenox-Conyngham, has also carried out important work, for which see its publications.

deformation, be k. Then the additional mass per unit area is $\rho' k$. Again, let the density at the highest level where plastic deformation has taken place be σ. Then compensation would be attained if the depression of the crust was $\rho' k/\sigma$, for this would make the mass expelled below equal to that added on top. If we neglect compressibility, this depression requires that the density after deformation at height x above a fixed equipotential surface within the layer of zero strength must be equal to the original density at a height $x + \rho' k/\sigma$, and is therefore $\rho + \frac{\rho' k}{\sigma}\frac{d\rho}{dx}$, where ρ is the original density at height x. The density diminishes towards the surface, so that the change in density at a given level due to the compensation, being $\frac{\rho' k}{\sigma}\frac{d\rho}{dx}$, is negative. This is constant if ρ is a linear function of x. Accordingly Hayford's hypothesis, that the defect of density is the same at all levels down to a given depth, and zero below that depth, is true on the hypothesis here adopted, provided that the density before deformation increases uniformly down to a given level, and is constant from that level down to the layer of flow. There is nothing inherently improbable about this, and it is clear that the charges of artificiality often made against Hayford's hypothesis are unfounded.

9·51. If there is a sudden change in density at any level, the compensation must be partly concentrated in that level. There is some evidence, from the behaviour of earthquake waves, that there is a discontinuity of substance at a depth of about 30 km., and the theory of the cooling of the earth already considered suggests that there must be a transition from lighter to denser rocks at a still smaller depth. Thus some of the compensation is probably concentrated at a depth of about 30 km., and probably an undue proportion of it is at still smaller depths. In addition there must be a gradual increase of density with depth, with a corresponding continuous distribution of compensation over a wide range in depth. Thus the true distribution of compensation would be expected to be intermediate between the uniform distribution of compensation, postulated by Hayford, and the concentration of compensation at one level required by the alternative hypothesis.

9·52. If Hayford's compensation is distributed through a depth H, the defect of density at any depth less than H due to an elevation of height h and density ρ' is $\rho' h/H$, and we shall have

$$h = k\left(1 - \frac{\rho'}{\sigma}\right).$$

Given the elevation of the land at all points of a country, and the density of the surface rocks, it therefore becomes possible to estimate the density anomaly at all points below the surface. From this, by means of an extremely laborious computation, in spite of the ingenuity of the officers

of the Survey in reducing this labour to a minimum, it is possible to calculate what the intensity of gravity and the deflexion of the plumb-line should be at each of a network of stations scattered over the United States, if Hayford's hypothesis is correct. The work is complicated by the fact that the hypothesis contains the unknown constant H, which is called the depth of compensation. It is therefore necessary to evaluate the anomalies on various hypotheses as to the value of H, and to compare the anomalies calculated for each value of H with those observed.

9·53. It will be seen that if H was infinite, the change of density would be zero at all depths, and there would be no disturbance of gravity other than that due to the elevation itself. Thus the case of infinite depth of compensation corresponds to the hypothesis of an undeformable earth. If H is zero, the compensation is all concentrated in the surface, so that the mountain has no attraction, and we have the Helmert hypothesis. Thus Hayford's hypothesis includes those of both Bouguer and Helmert as particular cases.

9·54. The following data, obtained by William Bowie*, summarize the residuals, on four different hypotheses, in the values of gravity at the Survey stations.

Mean residuals without regard to sign.

	$H=\infty$	$H=0$	$H=114$ km.	$H=60$ km.
Coast stations	0·021	0·022	0·018	0·012
Stations near coast (within 325 km.)	0·025	0·023	0·021	0·020
Stations in interior (not in mountainous regions)	0·033	0·020	0·019	0·019
Stations in mountainous regions, below general level	0·108	0·024	0·020	0·018
Stations in mountainous regions, above general level	0·111	0·059	0 017	0·022

Mean residuals with regard to sign.

	$H=\infty$	$H=0$	$H=114$ km.	$H=60$ km.
Coast stations	+ 0·017	+ 0·017	− 0·009	− 0·003
Stations near coast (within 325 km.)	+ 0·004	+ 0·017	− 0·001	+ 0·002
Stations in interior (not in mountainous regions)	− 0·028	+ 0·009	− 0 001	− 0·001
Stations in mountainous regions, below general level	− 0·107	− 0·008	− 0·003	0·000
Stations in mountainous regions, above general level	− 0·110	+ 0·058	+ 0 001	+ 0·016

The unit in each case is 1 cm./sec.^{2} (not 1 dyne as given by the original author).

In the first place, we see by inspection of these tables that no solution makes the mean of all the residuals without regard to sign less than 0·019 cm./sec.^{2} This is to be regarded as representing the irregular variation, and no mean residual can be considered significant unless it decidedly exceeds this standard. We notice then that no residual for coastal stations or stations within 325 km. of the coast can give useful information concerning the

* *Investigations of Gravity and Isostasy*, U.S. Coast and Geodetic Survey, 1917.

accuracies of the four hypotheses. In interior stations, not in mountainous regions, the Bouguer hypothesis begins to fail, giving a mean residual of 0·033 without, and − 0·028 with, regard to sign. In the mountainous regions, below the general level, the mean Bouguer anomaly has become 0·108 without, and − 0·107 with, regard to sign; and in mountainous regions, above the general level, it reaches 0·111 without, and − 0·110 with, regard to sign. In no case has the anomaly on either Hayford hypothesis risen above the ordinary limits of irregular variation. This is sufficient to make the Bouguer hypothesis untenable; the close agreement in absolute value between the residuals with and without regard to sign shows that the hypothesis is in error in the same sense at nearly every station. Some form of compensation must therefore be admitted.

The Helmert hypothesis is in sufficient agreement with the facts in the first four lines of each table, but breaks down completely when applied to the mountain stations, above the general level of the neighbourhood. As on the Bouguer hypothesis, the residual has the same sign at nearly every station; but the error is systematically in the opposite direction.

The failure of the hypotheses of the undeformable earth and of massless mountains to account for the observed variation of gravitation is conclusive evidence of the existence of some form of compensation within a finite depth. The compensation is not, however, necessarily complete, for residuals remain on all solutions; and it is possible that the depth of compensation, instead of being constant, is itself a function of position. Actually it is found that the residuals are made smallest for mountainous regions if the depth of compensation is taken to be 95 km., and for the less elevated regions if this depth is about 60 km.; but where the inequalities are small they could produce little disturbance of gravity even if they were uncompensated, and therefore little weight can be attached to a determination of the depth of compensation from them. The data are quite well represented by a uniform depth of compensation of 90 to 100 km.

Hayford, from the deflexions of the plumb-line, found the depth of compensation to be 97 km. Thus two entirely different sets of observations lead to almost identical results for the depth of compensation, a further confirmation of the hypothesis.

9·6. *Discussion of Residuals.* But though the geodetic data are enough to show that some kind of compensation exists, it is necessary to point out certain possible misinterpretations of the results. The Hayford hypothesis is not exact. If it were, all the residuals would be zero within the limits of error of a gravity determination, which at most stations does not exceed $\pm$ 0·003 cm./sec.^{2}; whereas the mean residual on either Hayford hypothesis is about 0·020 cm./sec.^{2} Though the hypothesis of complete compensation in a finite depth is a vast improvement on those of no compensation and of

compensation at zero depth, it is therefore not a complete statement of the facts. The theory can be complete only when the whole of the residuals have been reduced to the limits of observational error.

9·61. Two explanations of the residuals are readily suggested. The surface inequalities that give rise to them may be of extent not great in comparison with the depth of the region of weakness, so that the greatest stress-differences due to them will occur in the strong part of the crust, and only small ones in the weak part. Thus the depression required to reduce those in the underlying layers to less than the strength may be much too small to give compensation. The alternative is that the inequalities have a horizontal extent great enough to make great stress-differences occur at all depths within the strong part of the crust, but that part of the weight is supported by the stresses in the region of finite, but small, strength just below. These hypotheses may be compared by inspection of Figs. 11 and 12 of Bowie's memoir, giving the distribution over the United States of the residuals on the two Hayford hypotheses, the depth of compensation being taken as 114 and 60 km. in the two cases. The distribution follows much the same outlines in the two cases. The distance from one region of maximum positive residual to the next is in general of the order of 600 or 1000 km., so that the greatest stress-differences, if there was no plastic flow, would be at depths comparable with 100 or 160 km., and considerable stress-differences would exist at the depth where the crust becomes weak, seeing that strength disappears completely at about 400 km. The second hypothesis, that the imperfection of the compensation is due to the finite strength of the region of weakness, is therefore suggested by the general features of the distribution of residuals.

9·62. Let us consider what strength the residuals imply within the layer of weakness. If we suppose that a residual is due to an uncompensated hill, of height h and density ρ', we have seen that the gravity anomaly due to the attraction of the hill is the second term in the Bouguer formula, namely $\frac{3}{2}\frac{\rho'}{\rho_0}\frac{gh}{a}$. If we take

$$\frac{\rho'}{\rho_0} = \frac{1}{2}, \quad g = 1000\,\frac{\text{cm.}}{\text{sec.}^2}, \quad a = 6 \times 10^8\,\text{cm.},$$

we find that the attraction is equal to 0·020 cm./sec.^{2} if

$$h = 1{\cdot}6 \times 10^4\,\text{cm.} = 160 \text{ metres}.$$

The residuals therefore indicate that in general elevations or depressions of 160 metres are uncompensated. The vertical pressure due to such an elevation is $g\rho'h$, and the maximum stress difference is $\frac{2}{e}\,g\rho'h$, at a depth equal to $\frac{1}{2\pi}$ of the distance between successive crests. This amounts to

$3{\cdot}2 \times 10^7$ dynes/cm.^{2} The wave length of the inequalities being taken as 700 km., the greatest stress-difference in a perfectly elastic crust would be at a depth of 120 km., and the absence of compensation shows that strengths of the order of $3{\cdot}2 \times 10^7$ dynes/cm.^{2} must exist at that depth and at greater depths.

9·63. On the other hand, the greatest residuals anywhere in the United States are about 0·060 cm./sec.^{2}, if we except a station in the north-west where there is one of 0·093 or 0·100 according to the depth of compensation adopted*. Thus the greatest uncompensated inequality is about 480 metres in height. Since inequalities greater than this do not exist, it appears that the strength of the layer of weakness is not great enough to support an uncompensated inequality of this magnitude. This result implies that the strength of the layer of weakness is not greater than 10^8 dynes/cm.^{2}; for if a greater stress-difference could be borne it would have survived the adjustment that followed the great alterations in level that have taken place within the United States.

9·64. Taking the results of the last two paragraphs together, we see that between the depths of 100 km. and 400 km. there must be an extensive region where the strength lies between 3×10^7 and 10^8 dynes/cm.^{2} The crushing strength of basalt at the surface is $1{\cdot}2 \times 10^9$ dynes/cm.^{2} This great fall of strength between the surface and a depth much less than 400 km. is consistent with the fall that was expected from the theory of the cooling of the earth.

9·65. On the other hand, it has been seen that compensation implies more, not less, strength in the upper part of the crust than would be required to support uncompensated inequalities. A mountain range rising 10,000 metres above its adjacent valleys corresponds to a case where h is 5000 metres. This elevation, if possessed by a series of parallel ridges, would make the greatest stress-difference below them amount to 10^9 dynes/cm.^{2} This is nearly the crushing strength of basalt, which could therefore just support the Himalayas if the stresses required to support them could be distributed over an infinite depth. But the weakness below throws extra stress on to the upper part of the crust, and therefore the strength of the crust at a depth of 50 to 100 km. must be decidedly more than that of basalt at the earth's surface. This is in accordance with experimental evidence obtained by F. D. Adams and L. V. King, who have shown† that under the conditions of temperature and pressure existing at a depth of 18 km. Westerly granite acquires a strength of about 10^{10} dynes/cm.^{2} The available evidence, therefore, shows that the strength of the rocks of the earth's

* Further work by Bowie has shown that the anomalies near Seattle are due to abnormal densities at small depths, and not to a defect of compensation.

† *Journ. Geol.* 20, 1912, 97–138.

crust increases inwards until it becomes several times the strength of surface rocks; and that at some depth, probably of the order of 100 km., it commences to decrease, becoming small in comparison with that of surface rocks at a depth perhaps about 200 km., and finally disappearing at about 400 km. A variation of this type was first inferred by Prof. J. Barrell, in a series of papers in the *Journal of Geology* for 1914 and 1915. He applied the term 'asthenosphere' to the region of weakness where plastic deformation takes place under stresses distinctly less than the strength of surface rocks; the definition unavoidably is vague, since many different distributions of strength, within certain limits, would account equally well for the facts. He based his argument on the geodetic evidence; in the present work it has been based principally on the results of cosmogony, and the two methods of attack have been seen to give closely accordant results. Thus the facts of isostatic compensation, especially in relation to the depth of the region of weakness, are explicable by the theory of the origin of the solar system so far developed, and afford a striking confirmation of it.

9·7. *Widespread Inequalities.* It has been seen that an inequality whose horizontal extent is greater than 2000 km. will produce so much deformation in the layer of zero strength that it will be completely compensated. It follows that the continents, each taken as a whole, must be completely compensated. In other words, if we take average columns in a continent and in an ocean, their cross sections being equal, and their lower ends being on the same equipotential surface, the masses in the two columns will be equal. Thus the continents must be made of lighter materials than the ocean floors. It corresponds well with this that oceanic lavas are on the whole more basic, and therefore denser, than those of the continents; unfortunately knowledge of the nature of the plutonic rocks of the ocean floor is lacking, so that the comparison cannot be carried out for these. A quantitative observational test of this prediction has not yet been made. The tests of the theory of isostasy so far carried out refer almost wholly to differences of level within a continent, and not to the differences between continents and oceans. Hayford, in one of his computations, supposed the land inequalities uncompensated and the oceanic ones compensated, and found that the residuals were decidedly less than those of the Bouguer theory, showing that the ocean was probably compensated, at least within distances of 2000 km. or less from the coast. So far as this goes, it confirms the theory; but hitherto, except for the evidence of lavas, which is favourable, it has been found impossible to test directly the theory that the continents and oceans as a whole are compensated. The supposition stands, however, as a direct inference from a theory that has so far been amply confirmed, and will therefore be adopted in the following pages.

CHAPTER X

The Thermal Contraction Theory of Mountain Formation

"It were not best that we should all think alike; it is difference of opinion that makes horse races." MARK TWAIN, *Pudd'nhead Wilson.*

10·1. It is generally agreed among geologists that the principal cause of the elevation of mountains is that the crust of the earth is in a state of horizontal compression, under which it frequently gives way, the strata being then folded into a shorter length in the neighbourhood of the point where the crust has proved unable to withstand the stress-difference upon it. Such a compression seems to be the only mechanism that is qualitatively capable of producing folded mountains. Several possible causes of compression have been suggested, but the majority are inadequate to account for any appreciable fraction of the crumbling that has actually occurred. The most effective appear to be thermal contraction, which will be discussed in this chapter, and changes in the rotation of the earth, which will be dealt with later.

10·2. *Outline of the Theory.* It has been seen in Chapter VI that different parts of the interior of the earth have cooled since solidification by different amounts. In cooling they must have contracted in volume by different proportions, and in this way a state of stress must have been set up in the crust. The mathematical discussion of the character of the deformations produced was due originally to Dr C. Davison* and to Sir G. H. Darwin†. Consider the earth at some instant during its cooling, and consider the effect of the cooling that takes place during a further interval. Throughout the region from the centre of the earth to within about 700 km. of the surface, no appreciable change of temperature takes place, and therefore no change of volume. Between this level and the layer where cooling is most rapid, each layer cools more than the layer below it, and would therefore contract more if it were not obstructed by the matter below. The latter fixes the inner radius of this region, and therefore the requisite reduction in volume can be achieved only by reducing the outer radius. Thus the adjustment requires a thinning of this region without a corresponding reduction in its inner radius. Since this region is necessarily in the region of zero strength, the matter in it will adjust itself completely to the stresses involved, and assume a hydrostatic state. On the other hand, the outer surface of the earth undergoes no further cooling and contraction, and is therefore too large to fit the contracted region just considered. It will therefore be under a horizontal crushing stress.

* *Phil. Trans.* **178** A, 1887, 231–42.
† *Ibid.* **178** A, 1887, 242–49; or *Sci. Papers*, **4**, 354–61.

Since the region below the layer of greatest cooling becomes too small to fit the interior, while the outer surface becomes too large, there must be an intermediate layer where the contraction is just enough to enable it to continue to fit the interior. This layer is called the 'level of no strain.'

Below the level of no strain, the rocks at any time would, if they simply underwent a contraction in all directions in accordance with their cooling, be too small to fit the interior, and thus in order to fit it they are stretched out horizontally. Yield under horizontal extension may occur in two ways. If the rock is capable of flow, it merely spreads out horizontally and becomes thinner vertically. It may, however, fracture vertically, forming long fissures, which will descend to levels where flow is possible. In such a case, hydrostatic pressure will force the matter capable of flow up into the crack, and its further cooling will lead to its solidification or crystallization within the crack. In either case, the total volume of the matter within the level of no strain is the same, for the vertical crack cannot extend above this level, and therefore the motion of weak material into the crack only redistributes matter below the level of no strain without altering its quantity.

Above the level of no strain, it will be seen that the rocks have probably always been crystalline. Their strength must therefore be finite. Hence when exposed to horizontal stress they would not flow, but would behave in the manner of crystalline rocks under high pressure in the laboratory; they would fracture and bend up irregularly, starting at the weakest spot. Thus an elaborate system of folds would be produced in the upper rocks. These, according to the thermal contraction theory of mountain building, are the initial stages of mountain ranges.

Since there is no deformation initially, and cooling starts from the surface, the level of no strain must start from the surface and gradually descend. By the time it has reached any particular layer of material within the crust, that layer will be crystalline, and therefore will yield to subsequent deformation by folding. This folding will start at once, for the filling up of any fissures that may have been opened will have blocked them up with solid material; thus when compression is applied the rocks will not be free to move, and yield can take place only by crumpling. Accordingly the crumpling of any layer required to make it continue to fit the interior must be calculated from the time when the level of no strain passed that layer.

10·3. *The Amount of Compression available.* Let us now apply the above considerations in a quantitative discussion of the amount of crumpling that must have occurred on the earth, to enable the outside to continue to fit the interior. Consider a shell of internal radius r and thickness dr. Let its coefficient of linear expansion be n, where n may be variable, and let its initial density be ρ. Let the *rise* of temperature of

the shell be v; this will of course be negative. Then v is a function of r alone. The density of the shell, if we ignore the small change due to compressibility, becomes $\rho\,(1 - 3nv)$. Let the radius of the shell of radius r become $r\,(1 + \alpha)$. Then the external radius becomes

$$r\,(1+\alpha) + dr\left\{1 + \frac{\partial}{\partial r}(r\alpha)\right\} \qquad \text{.........(1).}$$

Hence the mass of the shell after the change of temperature is

$$4\pi r^2 (1+\alpha)^2\,dr\left\{1 + \frac{\partial}{\partial r}(r\alpha)\right\}\rho\,\{1 - 3nv\} = 4\pi\rho r^2 dr\left\{1 + 2\alpha + \frac{\partial}{\partial r}(r\alpha) - 3nv\right\} \qquad \text{.........(2),}$$

neglecting squares and products of α and v. But the mass is unaltered. Hence we have the equation of continuity

$$2\alpha + \frac{\partial}{\partial r}(r\alpha) - 3nv = 0 \qquad \text{.........(3).}$$

Since v is supposed known throughout the earth, this is a differential equation to determine α.

Now if a shell simply expanded without stretching, its radius would increase by rnv instead of $r\alpha$, so that the stretching required to make it continue to fit the interior is $r\,(\alpha - nv)$. Let us denote $\alpha - nv$ by k. Then substituting for α in (3) we have

$$\frac{d}{dr}(kr^3) = -\,r^3\frac{\partial}{\partial r}(nv) \qquad \text{.........(4).}$$

At the centre there will be simple expansion without stretching, so that k will vanish with r. Hence

$$k = -\frac{1}{r^3}\int_0^r r^3\frac{\partial}{\partial r}(nv)\,dr \qquad \text{.........(5).}$$

Let c be the radius of the earth. Then we can write

$$r^3 k = -\,[r^3 nv]_0^r + \int_0^r 3r^2 nv\,dr \qquad \text{.........(6),}$$

$$k = -\,nv + \frac{1}{r^3}\int_0^r 3r^2 nv\,dr \qquad \text{.........(7),}$$

since the integrated part vanishes at the lower limit. Now the change of temperature is appreciable only in a depth small in comparison with the radius of the earth. Hence in the region where k is appreciable r can to a first approximation be put equal to c. Then

$$k = -\,nv + \frac{3}{c}\int_0^r nv\,dr \qquad \text{.........(8).}$$

If we call the depth x, we have $r = c - x$ ....(9),

and then, since the temperature change is practically that in a solid of infinite depth with a plane face, as was seen in Chapter V, we have

$$k = -\,nv + \frac{3}{c}\int_x^\infty nv\,dx \qquad \text{.........(10).}$$

Now consider the changes that take place in a short interval of time dt. If the integral stretching in a particular layer since solidification be K, we have

$$k = \frac{\partial K}{\partial t} dt; \quad v = \frac{\partial V}{\partial t} dt \qquad \text{.........(11)},$$

where V is the temperature. Then (10) becomes

$$\frac{\partial K}{\partial t} = -n \frac{\partial V}{\partial t} + \frac{3}{c} \int_x^\infty n \frac{\partial V}{\partial t} dx \qquad \text{.........(12)}.$$

The stretching at the surface is to be found by putting x zero in this and integrating with regard to the time. The level of no strain is determined by the fact that $\partial K/\partial t$ vanishes there.

The coefficient of expansion is in general a function of the temperature, and can be approximately represented by

$$n = \epsilon + \epsilon' V \qquad \text{.........(13)},$$

where ϵ and ϵ' are two constants.

10·31. Considering first the case of a uniform distribution of radioactive matter, finite in depth, we had (6·6, equation 3) the approximate formula

$$V = mx + \left(S - \frac{AH^2}{2k}\right) \operatorname{Erf} \frac{x}{2ht^{\frac{1}{2}}} + \frac{AH^2}{2k} - \mu \qquad \text{.........(1)},$$

where $\mu = 0$ if x is greater than H, and $\mu = A\,(H - x)^2/2k$ if x is less than H. Putting $AH^2/2k = \alpha$, $S - \alpha = \beta$, $x/2ht^{\frac{1}{2}} = \lambda$, and evaluating 10·3 (12), we have

$$\begin{aligned} \frac{\partial K}{\partial t} &= (\epsilon + \epsilon'\alpha + 2\epsilon' m h t^{\frac{1}{2}} \lambda + \epsilon'\beta \operatorname{Erf} \lambda) \frac{\beta}{\sqrt{\pi}} \frac{\lambda}{t} e^{-\lambda^2} \\ &- \frac{3\beta h}{\sqrt{\pi t}\, c} \Big[(\epsilon + \epsilon'\alpha) e^{-\lambda^2} + \epsilon' m h t^{\frac{1}{2}} \{2\lambda e^{-\lambda^2} + \pi^{\frac{1}{2}} (1 - \operatorname{Erf} \lambda)\} \\ &+ \epsilon'\beta \left\{ e^{-\lambda^2} \operatorname{Erf} \lambda + \frac{1}{\sqrt{2}} (1 - \operatorname{Erf} \lambda\sqrt{2}) \right\} \Big] \qquad \text{.........(2)}, \end{aligned}$$

provided x is greater than H.

The level of no strain is found by putting $\partial K/\partial t = 0$, and solving (2) for λ. It is evident from the form of the equation that the appropriate value of λ is of the order of $3ht^{\frac{1}{2}}/c$, which is small. Hence we can approximate by neglecting λ^2; then (2) becomes

$$0 = (\epsilon + \epsilon'\alpha) \frac{\beta}{\sqrt{\pi}} \frac{\lambda}{t} - \frac{3\beta h}{\sqrt{\pi t}\, c} (\epsilon + \epsilon'\alpha + \epsilon' m h \sqrt{\pi t} + \epsilon'\beta/\sqrt{2}) \quad \text{......(3)},$$

which with the numerical data of 6·6, with Fizeau's values

$$\epsilon = 7 \times 10^{-6} \div 1^\circ \text{ C.},$$
$$\epsilon' = 2{\cdot}4 \times 10^{-8} \div (1^\circ \text{ C.})^2,$$

gives

$$\lambda = 0{\cdot}4 \qquad \text{.........(4)}.$$

The corresponding depth is 150 km. This is greater than H, which has

been seen to be 13 km., so that the omission of the terms depending on μ is correct. Thus the level of no strain is now below the layer of compensation.

To find the stretching at the surface, we put λ zero in (2). This assumes that the terms arising from μ can be neglected. They are evidently much less than that involving $\epsilon'\alpha$, which is itself less than the term in ϵ. This neglect is therefore justified. We have

$$\frac{\partial K}{\partial t} = -\frac{3\beta h}{\sqrt{\pi t}\,c}[(\epsilon+\epsilon'\alpha)+\epsilon' mh\sqrt{\pi t}+\epsilon'\beta/\sqrt{2}] \quad \text{.........(5)},$$

$$K = -\frac{6\beta}{\sqrt{\pi}c}\left[\left(\epsilon+\epsilon'\alpha+\epsilon'\frac{\beta}{\sqrt{2}}\right)ht^{\frac{1}{2}}+\tfrac{1}{2}\epsilon' mh^2\sqrt{\pi}t\right] \quad \text{.........(6)}$$

$$= -4{\cdot}1\times 10^{-3},$$

with the above data, for the stretching up to the present time. This estimate refers to continental regions. The corresponding estimate for oceanic regions would require an evaluation of the depth of the radioactive layer over the ocean floor, which cannot be done at present. If we assume that H is the same, but that A for the oceans is one-third of what it is for the continents, we find

$$\alpha = 57^\circ, \quad \beta = 1143^\circ,$$
$$K = -4{\cdot}5\times 10^{-3}.$$

The area of the land surface of the earth is $1{\cdot}45\times 10^{18}$ cm.^{2}, and that of the ocean surface is $3{\cdot}67\times 10^{18}$ cm.^{2} The relative reduction of an area being double that of a length, the surface of the continents has on this hypothesis been reduced by $11{\cdot}9\times 10^{15}$ cm.^{2}, and that of the oceans by $33{\cdot}1\times 10^{15}$ cm.^{2}, making a total of 45×10^{15} cm.^{2} This gives an estimate of the area of the rock surface that may have been crumpled up to form mountains.

10·32. Considering next the exponential distribution of radioactive matter, we had (6·61) the approximate formula

$$V = mx + \left(S - \frac{A}{a^2k}\right)\operatorname{Erf}\lambda + \frac{A}{a^2k}(1-e^{-ax}).$$

Substituting in 10·3 (12) and writing

$$A/a^2k = \alpha; \quad S - \alpha = \beta,$$

we find

$$\frac{\partial K}{\partial t} = \{\epsilon + 2\epsilon' mht^{\frac{1}{2}}\lambda + \epsilon'\beta\operatorname{Erf}\lambda + \epsilon'\alpha(1-e^{-2ah\lambda t^{\frac{1}{2}}})\}\frac{\beta}{\sqrt{\pi}}\frac{\lambda}{t}e^{-\lambda^2}$$
$$-\frac{3\beta h}{\sqrt{(\pi t)}\,c}\left[(\epsilon+\epsilon'\alpha)e^{-\lambda^2}+\epsilon' mht^{\frac{1}{2}}\{2\lambda e^{-\lambda^2}+\pi^{\frac{1}{2}}(1-\operatorname{Erf}\lambda)\}\right.$$
$$\left.+\epsilon'\beta\left\{e^{-\lambda^2}\operatorname{Erf}\lambda+\frac{1}{\sqrt{2}}(1-\operatorname{Erf}\lambda\sqrt{2})\right\}\right].$$

As before we see that where $\partial K/\partial t$ vanishes λ must be small, of order

$3ht^{\frac{1}{2}}/c$ or $\frac{1}{4}$. But $aht^{\frac{1}{2}} = 14$, and therefore $e^{-2ah\lambda t^{\frac{1}{2}}}$ can be neglected. Neglecting λ^2 elsewhere, we have

$$0 = (\epsilon + \epsilon'\alpha)\frac{\beta}{\sqrt{\pi}}\frac{\lambda}{t} - \frac{3\beta h}{\sqrt{\pi t}c}\left\{\epsilon + \epsilon'\alpha + \epsilon' mh\,(\pi t)^{\frac{1}{2}} + \frac{\epsilon'\beta}{\sqrt{2}}\right\},$$

to determine the level of no strain. This equation is exactly the same in form as 10·31 (3), but the values of α and β are different. We have, from 6·61, $\alpha = 340°$, $\beta = 860°$. Then $\lambda = 0{\cdot}3$, and the depth of the level of no strain is 111 km., just greater than the depth of compensation.

The stretching at the surface is given, as in 10·31, by

$$K = -\frac{6\beta}{\sqrt{\pi}c}\left[\left(\epsilon + \epsilon'\alpha + \frac{\epsilon'\beta}{\sqrt{2}}\right)ht^{\frac{1}{2}} + \tfrac{1}{2}\epsilon' mh^2\sqrt{\pi}t\right],$$

$$= -3{\cdot}6 \times 10^{-3}$$

for continental regions. For oceanic regions we had in 6·8

$$\alpha = 100°, \ \beta = 1100°.$$

This makes $$K = -4{\cdot}3 \times 10^{-3}.$$

The reduction of surface of the earth available for mountain building is therefore 42×10^{15} cm.², of which $10{\cdot}4 \times 10^{15}$ cm.² is derived from the continents and $31{\cdot}6 \times 10^{15}$ cm.² from the oceans. This result is slightly less than that found for the uniform distribution of radioactive matter down to a finite depth, but the difference is so small, considering the widely different character of the two distributions, that it appears very improbable that the actual distribution, whatever it may be, so long as it satisfies the conditions already shown to be necessary for any possible distribution, will give a value differing from this by more than a few per cent.

10·33. If we adopt the lower melting point of terrestrial rocks suggested in 6·7, namely

$$S = 800°,$$

we have, assuming an exponential distribution of radioactive matter,

$$\alpha = 330°;\ \beta = 470°;$$
$$K = -1{\cdot}6 \times 10^{-3},$$

for continental areas. For oceanic areas,

$$K = -2{\cdot}2 \times 10^{-3},$$

and the available compression is reduced to 21×10^{15} cm.², of which $4{\cdot}8 \times 10^{15}$ cm.² is derived from the continents and $16{\cdot}2 \times 10^{15}$ cm.² from the oceans. Thus the melting point assumed has a considerable influence on the available compression.

10·4. *Compression required to explain known Mountains.* It is necessary, in order to make a quantitative test of the thermal contraction theory of mountain formation, to compare the compression shown to be available for mountain formation on this theory with the compression required to produce known mountain ranges. A complete computation of the required

compression is not yet possible, on account of the incompleteness of the geological surveys of many of the greater mountain ranges. In the cases of a few ranges, however, geologists have succeeded in obtaining direct quantitative estimates of the amount of compression necessary to produce the observed folding. In the Appalachians, for instance, the width of the rocks, measured at right angles to the chain, has been shortened by about 40 miles. Similarly, the compression in the Rockies is 25 miles, in the Coast Range in California 10 miles, and in the Alps 74 miles*. The larger ranges in Asia and South America have not been so exhaustively treated; but a rough idea of their importance can be obtained by a comparison with the Rockies, or, in the case of narrower ranges, the Coast Range, whose geological age is about the same. The Alps are probably abnormal, and have not been used as a standard. The elevation of a continent or a large tableland involves little crumpling within it, and no great amount at the coast so long as the slope there is gradual. Hence in determining the amount of compression required, we need consider only the steep slopes of mountains. The compression at right angles to the range has been supposed proportional to the mean height, which would be exactly true if the strata were similarly folded in all mountains. In the following table the mean height has been obtained roughly from the maps

TABLE. *The great mountain ranges.*

	Range	Length (km.)	Mean height (metres)	Compression (km)	Area compressed (thousands of sq km)
Europe	Scandinavian	1400	1000	10	14
	Alps	1000	3000	118†	118
	Carpathians	1300	1000	10	13
	Apennines	900	700	10	9
	Urals	2200	700	10	22
	Pyrenees	400	2000	16	6
Asia	Caucasus and Armenia . .	3500	2000	40	140
	Iran	1400	2000	40	56
	Himalaya	4000	5000	100‡	400
	Suleiman, etc.	1200	2000	30	36
	Karakorum and Hindu Kush	2400	1000§	20	48
	Kwen Lun	2300	1000§	20	46
	Tian Shan	2000	3000	60	120
	Altai	1600	2500	50	80
Africa	Abyssinia	2000	2000	40	80
	Drakensberg	1500	1000	20	30
	Atlas	2000	1500	30	60
America	Coast Range	4000	2000	16†	40
	Rockies	7000	2000	40†	280
	Appalachians	800	1000	60†	48
	Andes	7000	2000	40	280
				Total	1926

* Pirsson and Schuchert, *Textbook of Geology*, 1915, p. 361.

† These are the data obtained directly from the geological evidence.

‡ This agrees roughly with a provisional estimate by Mr R. D. Oldham.

§ It is assumed that the folding needed to produce the plateau of Tibet was all at the margins, and is thus included in that found for the Himalayas and Tian Shan. Thus, for the mountains within the area it is only necessary to consider the height above the general level of the plateau.

in Philips' *Student's Atlas*. From this compression in width, with the length of the range, the area lost by folding is at once found. The amount of compression found is of course essentially provisional, and must be revised when further geological evidence is available.

It is thus found that in the formation of mountains the surface of the earth has been diminished by crumpling by about 20×10^{15} cm.^{2}

10·5. *The Pacific Mountains*. The available compression on the thermal contraction theory of mountain building has been seen to be over 40×10^{15} cm.^{2}, so that the amount available appears to be more than double what is required to account for existing mountains. The surplus, however, is not so great as this indicates. The compression of the ocean bottom is not all available for mountain formation. It might, in fact, be thought that none at all of it is so available, since it would give rise to mountains on the sea-bottom, and not to the observed continental mountains. This, however, is only partly true. On the present hypothesis the lower radioactivity of sub-oceanic rocks has enabled them to cool to a greater extent than sub-continental rocks. In addition, it appears that basic rocks are, on the whole, stronger than acidic ones; thus basalt has, under ordinary conditions, a crushing strength of $1{\cdot}2 \times 10^9$ dynes/cm.^{2}, as against 8×10^8 dynes/cm.^{2} for granite*. On both grounds, therefore, the rocks below the oceans must be stronger than those below the continents. Now where the compressed ocean floor abuts on a compressed continent, the weaker will be the first to give way; the continent margin will be forced inwards and its rocks piled up over those further inland. Thus ranges of mountains parallel to the coast will be formed, with much overthrusting. These correspond closely to the Pacific type of mountain, such as the Rockies and Andes. Such mountains are evidently, if this theory is correct, formed by the relief of sub-oceanic and not sub-continental compression, so that in these cases the compression produced has been derived from the oceanic rocks. Of the total compression required, about $6{\cdot}5 \times 10^{15}$ cm.^{2} was required to account for the Pacific type of mountains. If we assume that this has been derived from the compression of the ocean bottom, the compression required to account for other mountains is $13{\cdot}5 \times 10^{15}$ cm.^{2}, which is only slightly greater than the compression already seen to be available from the contraction of continental rocks alone. The theory is therefore certainly adequate to account for the greater part of the mountains known to us, and further has the recommendation that the existence of the two distinct types of mountain range known as the Continental and Pacific types is a direct inference from it. It is not impossible that, by a more complicated process, the continental type of mountains may have been partly formed by the relief of sub-oceanic compression, but this point will not be considered in the present work.

* Landolt and Bornstein, *Physikalische Tabelle*.

10·6. *Epochs of Diastrophism.* The formation of mountains has not taken place at all periods in the history of the earth; it is known that there have been long intervals of quiescence. These appear as a natural consequence of the present theory. In crustal rocks the modulus of rigidity is about 5×10^{11} dynes/cm.^{2}, and Young's modulus is about 12×10^{11} dynes/cm.^{2} The breaking stress being taken as 8×10^8 dynes/cm.^{2}, we see that if Hooke's law held right up to the breaking-point, the extension would be $-0{\cdot}7 \times 10^{-3}$. But it has been seen that the linear compression to be expected on the contraction theory is about 4×10^{-3}. Hence the compression at any place has had time to reach the breaking stress and undergo complete relief about six times.

What must happen on the thermal contraction theory of mountain formation is that the stresses increase continuously until the strength of the rocks is reached, when flow commences. Until this stage no flow and no mountain building occur, and we have an interval of quiescence. When the stress-differences reach the strength, complete fracture takes place in surface rocks, and the strength is reduced to zero. Thus crumpling continues until the stresses are almost completely relieved. This corresponds to an epoch of mountain formation. Then the fractures become sealed up afresh, and further internal cooling recommences the process. At any one place, by what has just been said about the strength of rocks, there may have been about six such epochs since the solidification of the earth. It is of interest that this is of about the order of the number of the great eras of mountain-building that are known to have occurred in the geological record.

10·7. It was thought by Osmond Fisher*, who has been followed by many geologists, that the thermal contraction theory of mountain formation is quantitatively insufficient to account for known mountain ranges. Osmond Fisher's discussion, however, rests on several incorrect hypotheses. In the first place, he uses Kelvin's theory of the cooling of the earth, which is now known to be in serious error, since it ignores radioactivity and takes the age of the earth to be only 10^8 years instead of over 10^9 years. This in itself shows that none of Fisher's quantitative estimates can be accepted at the present day. There is, however, a still more serious error. Fisher's criterion of quantitative accuracy is based, not on the reduction of area by crumpling, but on the volume of crumpled rock. This is estimated, correctly in principle, by finding the reduction of area in each layer down to the level of no strain, where crumpling ceases, and integrating with regard to the depth. Fisher then supposes the crumpled rock spread in a uniform layer over the surface of the earth, and finds that the thickness of this layer is only a few metres, which is small in comparison with the heights of existing mountains. With the data here employed, the depth

* *Physics of the Earth's Crust*, Macmillan, 1889.

found is greater than Fisher's estimate, being of the order of 200 metres, but is still less than the heights of known mountains. The comparison made, however, is quite illusory. It would be valid only if the physical process suggested in the calculation bore a close resemblance to that involved in the actual elevation of mountains, which is not the case. The rocks crumpled at great depths have not all been brought up to the surface in the process; a very small fraction of them have. Those crumpled at the surface have not been uniformly spread out, and if they had been, the surface would have remained perfectly level, and no mountains at all could have been formed. The only way of altering Fisher's comparison so as to make it serve as a trustworthy test of the thermal contraction theory of mountain building would be to find the depth of the layer that the existing mountains would form if they were all pulverized and spread uniformly over the earth; and it is certain that the depth of such a layer would not exceed a few metres.

10·8. It has been seen that the yielding of the crust below the level of most rapid cooling is an essential preliminary to mountain formation. Since this layer is in the region of small or zero strength, such yield would occur under very small stress-differences. The compressive movements of the strong part of the crust would follow this adjustment. Since the piling up of the crumpled rocks necessarily increases the weight of rock per unit area locally, enough flow must take place below to reduce the stress-differences at all levels to the strength. Hence if the folding takes place over a great enough area, the mountains will be isostatically compensated from their birth. The formation of mountains and their isostatic compensation are both parts of the same process.

10·9. *Summary*. It has been shown that the thermal contraction theory of mountain building, with the data already adopted from other evidence, implies a reduction of the surface of the earth by crumpling of the order of 4×10^{16} cm.^{2}, while the reduction required to account for known mountains is about 2×10^{16} cm.^{2} On the whole the thermal contraction theory appears able to account for the greater part, and possibly the whole, of the existing mountain ranges. It easily explains, in addition, the difference between the Pacific and Continental types of mountain range, and the long intervals of quiescence between the great epochs of mountain formation. Osmond Fisher's argument against it appears to be fundamentally fallacious. The great mountain ranges have probably been isostatically compensated from their formation.

CHAPTER XI

Theories of other Surface Features

"The time has come, the Walrus said,
To talk of many things."

LEWIS CARROLL, *Through the Looking-Glass.*

11·1. *The Origin of Oceanic Deeps.* In consequence of the greater cooling at depths of 100 to 200 km. below the oceans, the rocks there must have tended to contract more than rocks at equal depths below the continents. The surface, on the other hand, must have remained at almost constant temperature on account of the cold water at the sea-bottom. This would give rise to a phenomenon comparable with the bending of the covers of a book when held in front of a fire. In the latter case the contraction of the side nearest to the fire is due to loss of moisture, whereas in the case of sub-oceanic rocks it is due to cooling; but the contraction is an essential feature of both phenomena, and its effects must be similarly shown in a tendency to curl. In each case the margins bend towards the contracting side; thus the ocean floor will tend to sink near the coasts and to rise in the middle. In the case of the book-cover there is nothing to oppose the curling, and complete adjustment is therefore possible; but in the earth's crust the tendency is resisted by the upward pressure of the underlying rocks, which has to be overcome in the depression of any portion of the crust. The tendency to force the asthenosphere downwards at the margins and upwards at the centre creates a stress-difference in the asthenosphere, which will yield so as to make this small again, the crust meanwhile bending so as to fit the new form of the asthenosphere.

At a depth of the order of 100 km. oceanic rocks must have cooled about 300° more than continental ones. Young's modulus for materials at this depth is known from earthquake data to be about 15×10^{11} dynes/cm.^{2}, and hence the tension necessary to prevent contraction, with the coefficients of expansion already adopted, must be about 8×10^9 dynes/cm.^{2} The crushing strengths of ordinary rocks at atmospheric pressure are much less than this, but it has been seen that even below the continents the rocks at such depths as 100 km. are probably much stronger than at the surface, and that sub-oceanic rocks are probably stronger still. Hence the rocks below the ocean floor may be able to withstand the whole of this tension, and we may proceed to discuss the behaviour of the ocean floor on the hypothesis that there is no yield.

The determination of a complete formal solution of the problem of the straining of a portion of the earth's crust by cooling would be very laborious. As only an estimate of the order of magnitude of the possible

deformation is required, we shall, therefore, discuss only the following related problem for the cylindrical case.

A fluid is in equilibrium in the form of an infinite circular cylinder, the force of gravity at the surface being normal to the axis. On its surface an infinite strip of solid matter floats, its curvature being equal to that of the surface of the fluid. It then contracts by cooling, the relative contraction that would be produced at any depth, x, in the absence of stress, being α. The solid is supposed to have Young's modulus E, and Poisson's ratio zero. It is assumed that after contraction it remains a portion of a cylinder. Find the elevation of the centre above the margins.

Let $2R$ be the width of the strip, c the radius of the cylinder and strip before cooling, and c_1 that of the strip after cooling. Let the elevation of any point of the surface above the undisturbed position be ζ, and the difference between the values of ζ at the centre and at the margins, H. Then the original length of the arc at depth x was $2R\left(1-\frac{x}{c}\right)$, and the natural length after cooling is $2R\left(1-\frac{x}{c}-\alpha\right)$. The actual length is $2R\left(1-\frac{x}{c_1}\right)$. Hence the amount of stretching is $2R\left(\frac{x}{c}+\alpha-\frac{x}{c_1}\right)$, and the tension is $E\left(\frac{x}{c}+\alpha-\frac{x}{c_1}\right)$. The strain energy is therefore $ER\int\left(\frac{x}{c}-\frac{x}{c_1}+\alpha\right)^2 dx$ taken through the strip. The gravitational energy is $\int\frac{1}{2}g\rho\zeta^2 dy$, where y is the element of arc of a section of the cylinder, and the integral is taken across the strip. The approximate solution desired will be obtained by making the total energy a minimum.

Remembering that the mean value of ζ must be zero, from the condition that the strip is still floating, we see that

$$\zeta=\left(\frac{1}{3}-\frac{y^2}{R^2}\right)H \qquad \text{.........(1).}$$

We also notice that

$$c_1\left(1-\cos\frac{R}{c_1}\right)=c\left(1-\cos\frac{R}{c}\right)+H \qquad \text{.........(2).}$$

If powers of R/c above the second be omitted, this gives

$$1/c_1-1/c=2H/R^2 \qquad \text{.........(3).}$$

The energy reduces to

$$\frac{4}{45}g\rho H^2R+ER\int\left(\alpha-\frac{2Hx}{R^2}\right)^2 dx \qquad \text{.........(4).}$$

The equilibrium condition is obtained by differentiating this with regard to H. Hence

$$\frac{8}{45}g\rho H-E\int\frac{4x}{R^2}\left(\alpha-\frac{2Hx}{R^2}\right)dx=0 \qquad \text{.........(5).}$$

With $g=981$ cm./sec.^{2},

$$\alpha=4\times10^{-3},\quad x=100\text{ km.},\quad R=2000\text{ km.},$$

we see that the first term is of the order of sixty times that under the integral sign involving H. In other words, the tendency to curve is mostly balanced by the disturbance of hydrostatic pressure, so that the curvature that actually takes place is only a small fraction of what would occur in the absence of gravity. With the above values of the quantities involved, H is found to be nearly 1 km. Thus, we should expect to find that the chief oceans will have regions around their margins deeper, by a distance of the order of a kilometre, than the centres.

This bears a suggestive resemblance to the facts with regard to oceanic deeps. All the chief deeps in the Pacific are near the margin: there are depths of 8500 m. near the Kurile Islands, 6000 m. off the Aleutian Islands, 7600 m. off Chile, 9000 m. near the Marianne Islands, and a strip of depth 9000 m. to the north of New Zealand. The last two are now in mid-ocean, but are near the edge of a former continent. The mean depth of the oceans is about 5000 m., so that some of these regions are too deep to be altogether accounted for in this way unless the difference of cooling is supposed to extend to a greater depth than 100 km.: but this, of course, is quite probable. The same is true on a smaller scale in the Indian Ocean. Off the coasts of Australia, Java and Sumatra there are depths of 6000 m., while the middle is occupied by a huge area whose depth does not exceed 4000 m. In the Atlantic, again, shallow strips extend from the south up the middle, past S. Georgia, Tristan d'Acunha, St Helena, and Ascension, and from the north right down the centre to the Azores. Thus each ocean has a region of smaller depth in the middle, as is predicted by the theory.

A possible test of this theory is afforded by the fact that the inflow of matter from the margins towards the centre required by the theory is not caused by any additional weight on the surface. It therefore corresponds to a net transference of mass towards a particular region. The mass per unit area at a deep should therefore be less than that at the centre of the ocean. This should be indicated by a true defect of gravity at sea-level in the deeps; in a gravity determination it should therefore appear as if the deeps were uncompensated. The evidence at present available on this point is meagre. Duffield's observations* on the *Morea* indicate a defect of gravity over the deep parts of the Indian Ocean of about the theoretical amount, but are open to some uncertainty. Nevertheless it would be a remarkable coincidence if accidental errors should have happened to produce low values of gravity at all the places where they should theoretically have been expected, and the observations, so far as they go, definitely support the theory.

The gravity anomalies here predicted do not arise from a permanent strength in the asthenosphere, which is on this theory in a hydrostatic state, as under a mountain chain; they arise from the strength and tendency to

* *On the Determination of Gravity at Sea*, Brit. Assoc. Report, Newcastle, 1916, pp. 549–65.

curvature of the strong upper crust. Faulting may occur near the margins. A reversed effect may be looked for in the continents, but is probably masked by denudation and sedimentary rocks.

11·2. *The Formation of Geosynclines.* It has been seen (9·52) that the addition of a thickness h of matter of density σ depresses the crust by an amount $\sigma h/\rho_0$, so that the upper surface of the new matter is $\left(1-\frac{\sigma}{\rho_0}\right)h$ above the original surface. If the deposition takes place from water, the additional mass is only the excess of the mass of the sediments over that of the water displaced. Let ρ_1 be the density of the sediments and ρ_2 that of the water, h the depth of sediment deposited, and x the depression of the original surface. Then the mass per unit area is increased by

$$\rho_1 h - (h-x)\,\rho_2 - \rho_0 x,$$

and the condition for compensation is that this shall vanish. We find that

$$(h-x)/h = (\rho_0-\rho_1)/(\rho_0-\rho_2) \qquad \text{.........(1).}$$

But $h-x$ is the reduction in the depth of the water. We therefore see that the depth of sediment that can be deposited in a sea whose initial depth is known is a determinate multiple of this depth. If we take ρ_0 to be 3·2, which is probably typical of the rocks at a depth of some hundreds of kilometres, ρ_1 to be 2·2, and ρ_2 to be 1·0, it is seen that the maximum depth of the sediments is 2·2 times the original depth of the water.

The possibility of deposition of sediments to a depth far greater than the initial depth of the water in which they are formed is obviously of considerable geological importance; but the depression of the sea-bottom that can be produced in this way has been much exaggerated, as has been pointed out by A. Morley Davies*. If the stress-differences present become too small to cause flow, compensation will not proceed. Accordingly, the adjustment that takes place can never be greater, and may be less, than the amount needed to give compensation. But very many cases are known where the existing deposits are far more than 2·2 times as deep as the sea can have been when their formation commenced, and for these the theory of isostatic compensation is therefore inadequate. To account for them we must have a theory that will explain how the crust in a region of deposition can be depressed by an amount far in excess of that needed to give compensation. Again, some of these sedimented regions afterwards rise far above sea-level, implying a flow of matter into them for which there could be no explanation if the depression of the crust at the end of the sedimentation was less than or equal to the amount needed to neutralize the effect of the weight of the sediments. Accordingly, there must be important departures from isostasy at certain stages of the development of such regions, and no explanation of the existence of sedimentary rocks above sea-level can be satisfactory unless it takes them into account.

* *Geological Magazine*, 1918, pp. 125 and 233; E. M. Anderson, *loc. cit.* p. 192.

A method by which these extensive sedimented regions can afterwards be uplifted is suggested by the theory of the origin of deeps just described. The sediments from a continent must often be deposited in a gradually developing deep. Their weight will accentuate the tendency already existing for that region to sink. Hence the stresses in the crust will be increased, and may lead to fracture. When this takes place, the crust on the oceanward side will be free to bend down further. That on the landward side of the fault, however, will now have nothing to hold it down except the weight of the sediments. Accordingly, it will be free to rise above sea-level. If compression occurs afterwards, a greater thickness of light sedimentary rocks is accumulated, and hence the surface will be raised still higher.

11·3. *The Stresses in a Cooling Earth before Set has occurred.* If the difference between continents and oceans be ignored, and the earth be considered simply as cooling from the outer surface, let us consider the elastic strain due to the change of temperature. The earth in these conditions will remain perfectly symmetrical, and if the centre be taken as the origin of coordinates, we have, with the notation of Chapter VII,

$$u = qx/r, \quad v = qy/r, \quad w = qz/r \qquad \text{.........(1)},$$

where q is the radial displacement and r the distance from the centre.

The radial force acting on unit mass is g, and is of course in general a function of r. Thus

$$X_0 = -gx/r, \text{ etc.} \qquad \text{.........(2)},$$

and

$$uX_0 + vY_0 + wZ_0 = -gq \qquad \text{.........(3)}.$$

The force acting on the same particle after the displacement is

$$X_0\left(\frac{r-q}{r}\right)^2, \text{ etc.,}$$

so that

$$X_1 = 2gxq/r^2 \qquad \text{.........(4)}.$$

Also

$$\delta = \frac{1}{r^2}\frac{d}{dr}(r^2 q) \qquad \text{.........(5)},$$

$$\rho_1 = -\rho_0\delta - q\frac{d\rho_0}{dr} \qquad \text{.........(6)}.$$

Substituting in 7·1 (22), and remembering that λ and μ are functions of r alone, we find that all three equations are satisfied if

$$\frac{d}{dr}\{(\lambda + 2\mu)\delta\} - \frac{4q}{r}\frac{d\mu}{dr} - \frac{d\gamma}{dr} + \frac{4q}{r}g\rho_0 - \rho_0 q\frac{dg}{dr} = 0 \qquad \text{.........(7)}.$$

If H be of the order of magnitude of the depth to which cooling has extended, we see that in the portion of the earth affected by the cooling dq/dr must be of order q/H, and therefore in general large compared with q/r. If c be the radius of the earth, we see that the first term in this equation is of order $\lambda q/H^2$, the second $\lambda q/c^2$, and the fourth and fifth $g\rho_0 q/c$. Accordingly no term, with the exception of the third, can amount to more

than a hundredth of the first. We can accordingly reduce the equation to

$$\frac{d}{dr}\{(\lambda + 2\mu)\,\delta\} - \frac{d\gamma}{dr} = 0 \qquad \text{.........(8)},$$

giving

$$\delta = \frac{\gamma}{\lambda + 2\mu} + \text{const.} \qquad \text{.........(9)}.$$

If we consider a point on the axis of x, the additional stresses are

$$p_{xx} = \lambda\delta + 2\mu\frac{dq}{dr} - \gamma \qquad \text{.........(10)},$$

$$p_{yy} = p_{zz} = \lambda\delta + 2\mu\frac{q}{r} - \gamma \qquad \text{.........(11)},$$

$$p_{yz} = p_{xz} = p_{xy} = 0 \qquad \text{.........(12)}.$$

Thus the principal stresses are radial and tangential, and it is at once seen from symmetry that this must be true at all points. Now the tendency of the matter to flow or fracture is determined by the stress-difference. As the initial stress-difference was zero, the actual stress-difference is the same as the difference between the radial and tangential additional stresses. If P, the radial stress, is the greater, horizontal fracture will tend to occur; if Q be the greater, the fractures will be vertical. Now

$$P - Q = 2\mu\left(\frac{dq}{dr} - \frac{q}{r}\right) = 2\mu\left(\delta - \frac{3}{r^3}\int_0^r r^2\delta\,dr\right)$$

$$= \frac{2\mu}{r^3}\int_0^r r^3\frac{d\delta}{dr}dr = \frac{2\mu}{r^3}\int_0^r r^3\frac{d}{dr}\frac{(3\lambda + 2\mu)\,nV}{\lambda + 2\mu}dr \qquad \text{.........(13)}.$$

Put

$$\frac{(3\lambda + 2\mu)\,nV}{\lambda + 2\mu} = \theta.$$

Then

$$P - Q = \frac{2\mu}{r^3}\int_0^r r^3\frac{d\theta}{dr}dr \qquad \text{.........(14)}.$$

Suppose first that the last adjustment to stress took place at solidification. Then the cooling is greatest at the surface, and steadily becomes less inwards. If λ/μ is nearly constant, as is usually true, $d\theta/dr$ is always negative, and therefore $P - Q$ is always negative. Thus the immediate effect of the cooling of the earth is to produce a strong tendency to vertical fracture at all depths.

On the other hand, suppose that the crust has by flow or fracture adjusted itself since solidification until the horizontal tension first produced has been completely relieved, and consider the effect of further cooling. The fall of temperature at the surface is zero, on the supposition that the temperature there is maintained wholly by radiation from the sun, which is supposed constant. Hence θ is zero when $r = 0$, falls to a maximum negative value in the crust, and then increases again, reaching zero again at the surface. Thus if it were not for the variation of r^3 within the region of integration, $P - Q$ would be zero at the surface. But above the layer of greatest cooling r is greater than below, and therefore in the integral

$d\theta/dr$ is multiplied by a larger quantity when it is positive than when it is negative. Thus $P - Q$ is positive when $r = c$, and of order $c^2H\theta_1$, where θ is the numerically greatest value of θ. At depths comparable with that of the level of greatest cooling, the integral is of course negative. Hence if there has been no variation in the surface temperature since hydrostatic conditions were last attained, symmetrical cooling must necessarily lead to a horizontal compression at the surface and a tension below.

11·4. *The Origin of Terrestrial Continents and Lunar Maria.* It has just been shown that the cooling that immediately followed solidification must have produced a tremendous tension in the uppermost layers of the crust. This tension would be practically that which would be developed if a rock cooled down from near its melting point to ordinary temperatures while its ends were kept immovable. No rock could stand such a tension, and, accordingly, it must soon have been relieved in some way. The nature of the relief requires discussion; its effects appear to have been neglected by geologists, presumably because it took place before the oldest known rocks were formed, but, nevertheless, it is likely to have played a very important part in determining the present configuration of the earth's surface; and relief of tension in modified form has probably continued to produce notable effects even up to the present day. Sir G. H. Darwin, in his investigation of the amount of mountain building to be expected on the Kelvin theory of the cooling of the earth, seems to have thought that the relief would take place by horizontal flow, the surface layers merely becoming somewhat thinner without change of length, and thereby acquiring a new unstressed state. This may be a satisfactory description of the phenomena at great depths, where the pressure is great; but at the surface a rock under tension would break at right angles to the tension just as any rope or bar does in air. Accordingly, the surface must have become honeycombed with vertical cracks. The depths of these would initially be very small and nearly equal, but the differences in depth would gradually grow. For suppose that a crack, A, is slightly deeper than a neighbouring one, B. Then further cooling below will produce a new tension, and the crust at A will have been more weakened by the deeper crack there, and therefore the crack A will commence to grow downwards sooner and more rapidly than the other. When cooling has progressed a long way down, the cracks must become very unequal in depth, and only a few of those originally formed will then be deep enough to continue their growth.

Now it must be remembered that, before solidification, the temperature was not uniform, because the melting point would be raised by pressure, and would therefore rise with the depth. Hence a crack extending downwards must be penetrating regions of higher and higher initial temperature as it proceeds; but its internal pressure is necessarily atmospheric and

practically constant. Hence, although the rocks at any depth are necessarily below their melting points at the pressure normally appropriate to that depth, as soon as they are reached by a crack, there will be a fall of pressure which may lower the melting point sufficiently to cause fusion; all that is needed is that the crack may reach a depth where the actual temperature is as high as the initial temperature at the surface. It may be easily shown from the equation 11·3 (13) that this will be achieved at the depth of most rapid cooling, which is also a region of great tension, when cooling has proceeded for an interval of the order of 10^7 years, when the depth of the cracks would be comparable with 8 km. When this happens, fusion must take place, and magma will be forced up the crack by hydrostatic pressure until the horizontal uniformity of pressure is nearly restored. Now the density of the matter there was probably not very different from that at the surface, and the semi-fluid magma may even have been lighter than the solids at the surface. Hence hydrostatic conditions would not be restored till the crack was practically full. Thus intrusions, not unlike the dykes of the present day, would be formed. Known dykes are not, of course, original examples of this process, being of much later date; all sign of these primitive dykes must have been buried beneath sediments and igneous outpourings long ago. On the moon, however, denudation and sedimentation do not exist, and there some relics may be sought. The well-known rills are not instances, being of later date, as is seen from the fact that in some cases they have broken through crater walls. The radiating streaks are much more likely to afford examples. These are narrow streaks, radiating as a rule from large craters; that they are filled up to the level of the surface is plain from the fact that they are extremely difficult to see under oblique illumination, which would not be the case if there were any difference of level that could cause a shadow to be thrown. In fact, the agreement in level is surprisingly good, considering that it can only arise from a more or less accidental numerical coincidence between the average density of the rocks down to the bottom of the crack, and the density of the rocks at the bottom when fused. The fact that the theory calls for such a coincidence does not, however, afford any argument against it, for the extreme smoothness of the surface of these streaks shows that they must have been filled with a semi-fluid at one time, and the support of this would have to be explained by some such balance, whatever theory we should choose to adopt as an explanation of their origin. So far, therefore, we may hold that the theory is confirmed by the existence of these streaks on the moon. As it depends partly on the increase of pressure with depth, which would be greater in the earth on account of the greater value of gravity, we may have some confidence in its applicability to the earth.

When the lateral ends of cracks are near together, the short length of crust between them has to support the whole of the unrelieved tension, and is therefore a particularly likely part of the crust for the next fracture to

occur. Cracks will therefore tend to grow together, and thus will tend to develop into closed polygonal systems. When this takes place, a qualitatively different stage of the process commences. Each polygon is separate from the rest of the crust, and will therefore proceed to develop on its own account. Its surface has long ceased to change in temperature, but cooling continues below, so that there is a tendency to contract underneath. This would tend to close the cracks above by shortening the crust, were it not that the cracks have been closed already by the injected magma. Hence the shortening below can be achieved only by curvature of the crust. The centre of the polygon must therefore rise in the middle relative to the edges; its centre of gravity cannot rise, however, since that would imply the existence of a great additional pressure over the surface, which the weak matter just below would be unable to support. Accordingly, while the centre must rise, the boundary must sink. The matter below will offer little resistance to this depression, but rather will make way for some of it by breaking through the dykes that form the boundary. What reaches the surface will fuse, owing to the relief of pressure, and flow out so as to submerge the depressed portion. It is obvious from hydrostatics that it must rise to a level above the tops of the cracks, for a simple fracture would bring it nearly level with them, even if the margins were not bent down, and the curvature would be enough to send them far below the surface. When the ejected matter solidified, which would not take long, a smooth surface would be formed. Here, again, we find a verification on the moon, for the large maria are extensive regions of great smoothness, and the regions between them are at higher levels. The bounding cracks would of course have been submerged below the outpour and become lost to sight for all time. What is particularly interesting about the maria, however, is that all the chief lie in a chain, forming the greater part of a circle, about 1000 miles in diameter, so that the suggested formation of a raised polygon is confirmed. The fact that they are darker than the average of the lunar surface, while the streaks are brighter, may perhaps be attributed to the different conditions of solidification inside a crack and in the open, or perhaps to the matter having come from a different depth.

The submersion of the matter around the edges of these polygons below hotter matter, far above its normal melting point, must have caused the depressed rocks there to melt again, or at least to soften. Now the density, by hypothesis, increases with depth, and therefore the melted matter, being originally derived from a higher level in the crust, is lighter than its surroundings and tends to rise. The highest part of the polygon being the middle, on account of the curvature, the fused material would collect there. Thus the centre would become characterized by a greater depth of light matter than the edges; but if the average density down to a certain equipotential is excessively low at a place, a greater depth is required to give

compensation. Hence, when the fused matter solidified again, even if the crust afterwards gave way under the tension involved, the centre would still remain elevated above the margins. Thus a permanent departure from sphericity would be produced. The region on the moon that has apparently been uplifted in this way is about 1000 miles across; a region on the earth of the same relative size would have a diameter of 3600 miles, about the width of Africa or North America. If such a process ever took place on the earth, we should therefore expect it to have led to the formation of elevated regions of similar size to our actual continents.

Before developing this hypothesis further, let us consider certain other data about the moon which may be relevant. The lunar craters are the dominant physical feature of the continents, but they are almost absent from the maria. This is probably due to their having been formed before the maria; most of those that were originally present would then have been submerged in the outflowing lava and hidden, and only those of later origin would have examples in the maria. Streaks also are rare in the maria, for the same reason. It may be objected to the theory that after the outflow all tension in the crust would be relieved, and that therefore no cracks at all could be formed subsequently. The partial fusion of the surface by the hot liquid must, however, lead to the formation of a new hot solid surface, when tension could begin afresh, though with less violence than before. The origin of craters I have not attempted to account for. The rills may be analogous to rift valleys on the earth.

The origin of continents offers one of the most difficult problems in geophysics. Numerous attempts have been made to solve it, but none of the theories offered appears satisfactory. The tetrahedral theory is one of the best known; this starts with a newly solidified earth, and it is supposed that the contraction of the inner parts left the outer crust in a state in which it retained its original area, but had to collapse so as to accommodate itself to the reduced volume of the interior. The form adopted would, it was said, differ from the sphere in the direction of the regular solid with largest surface for the given volume, namely, the regular tetrahedron, and thus four continents would be formed, all at equal distances from one another. The physical aspect of the theory has not been considered in detail. It derives some support from the fact that the actual distribution of land does bear some resemblance to a tetrahedron, though this is probably a peculiarity of the present time, and seems to have been widely departed from at some previous epochs. The fatal defect of the theory, it seems to me, is that a tetrahedral deformation does not correspond to a figure of equilibrium. It is known that for any such displacement the elevated parts must tend to come down again, since both gravity and the curvature of the elastic outer layer act so as to restore the original state. Instead of retaining the deformed figure, the earth would therefore oscillate about the spherical form till the oscillations were damped out,

when symmetry would be restored. The shell, being too large for the interior, would then be unsupported and would collapse. A tetrahedral deformation cannot therefore be produced in this way; the way to render one possible is to have the continents free at their edges, so that curvature can take place in consequence of the natural cooling. The curved continent will then practically float on the heated interior, and any oscillation that may take place will merely move it up and down with the interior. Thus the process we have indicated is an essential preliminary of a tetrahedral deformation.

It has at various times been suggested, especially by Osmond Fisher, that the birth of the moon may have had an important effect on the distribution of land and sea on the earth. This view must be examined with special care, because the hypothesis that the moon was formed from the earth by the disruptive action of the solar tides has acquired a considerable probability, as has been seen in Chapter III, though it cannot be regarded as demonstrated.

It was shown in 5·83 that the moon, if formed in this way, must have been formed when the earth was almost fluid, with at most a thin solid crust on the outside. In the violent agitation that took place during the process, this thin crust must have been broken into fragments, floating on the liquid interior. If they stayed where they were, the removal of a large quantity of surface matter from one side would leave a vast area with no light matter, which would in course of time develop into the present Pacific Ocean. Unfortunately, however, the fragments would not stay where they were. Light solid bodies floating on a liquid interior, and largely confined to one side, would correspond to a first harmonic deformation, which has been shown to be unstable*. They would spread themselves out in such a way as to get as far apart as possible, thus allowing the denser liquid below them to get as close together as possible, and thus would become roughly symmetrically distributed over the earth at once; the scar left by the birth of the moon would have lasted a few days or months instead of a thousand million years.

A further criticism of the theory is that even if the Pacific Ocean could be accounted for in this way, a similar explanation would not be available for the asymmetry of Mars; for Mars has certainly not acquired a satellite by fission, and, even if it had, the equator would run through the middle of the ocean produced, whereas the dark regions on Mars, which are widely believed to be seas, are nearly confined to the southern hemisphere. There must therefore be some alternative reason for an asymmetrical distribution of oceans.

11·5. *Effects of Decrease of Density within the Crust.* In the theory of the cooling of the earth, which has been utilized to account for the de-

* J. H. Jeans, 'Gravitational Instability and the Figure of the Earth,' *Proc. Roy. Soc.* 93 A, 1917, 413–17. The argument of 13·52 amounts to an alternative proof.

velopment of the principal surface features, it has consistently been supposed that every element of the crust has fallen steadily in temperature, and that every element decreases in volume as it cools. The first of these suppositions would be strictly correct if the radioactive matter of the crust were symmetrically distributed, the rate of generation of heat per unit volume being a function of the depth alone, and independent of the position with regard to the earth's surface. It has been seen, however, that the difference between continental and oceanic conditions points to an asymmetrical distribution of radioactive matter. This in itself does not indicate any possibility of internal heating of the crust, since the original calculation referred to the continents, and the evidence indicates that oceanic rocks develop less internal heat than continental ones. If, however, for any reason the radioactivity in some locality is decidedly greater than that in typical continental regions, heat may be developed locally in excess of the loss by conduction, and the temperature may rise. This may occur especially in cases of mountain formation by crumpling. The shortening of the crust has already been seen to imply a concentration of granitic rocks in the region of upheaval, and these rocks represent the most radioactive type. It may therefore be expected that local heating may take place below mountain ranges. This is confirmed by the fact that the rate of increase of temperature with depth observed in high mountains is considerably above normal. The addition of sediments to a continent margin, again, must tend to increase the internal temperature; though this process is not likely to be so potent in this respect as mountain formation, on account of the lower radioactivity of the added matter.

The second supposition, that the coefficient of thermal expansion is positive, is also true in most cases. It is possible, however, that some rocks may expand in solidification, or in the transition from the liquevitreous to the crystalline state. If so, a temporary expansion will be an incident of the cooling process.

Thus there may be local and temporary exceptions to the general rule that the matter of the earth's crust has steadily contracted since solidification in general accordance with the theory of Chapters VI and X. Let us then consider the effects of any local expansion that may have occurred. The expanded material must become too large for its surroundings; this will indeed happen even when contraction takes place everywhere, provided that it is less in some restricted region than in the surrounding parts. Thus a state of stress will be set up. Evidently the principal stresses will consist of pressures across the boundary of the region and tensions parallel to it. Hence there will be a tendency for cracks to develop across the boundary and for the expanded matter to be forced into them. If the stresses become great enough (a possibility that has not yet been quantitatively tested) such a rupture and intrusion will occur. Whether the cracks will reach the surface will depend on the special circumstances of the case.

On this theory the great mountain ranges are the most probable regions for intrusion, since they are necessarily places where the underlying rocks have cooled less than elsewhere. Intrusion cannot, however, reach the surface until a horizontal tension has been developed: if there were a horizontal pressure it would close the cracks as fast as they were formed—in other words, it would prevent their formation. The original horizontal pressure that produced the mountain ranges is relieved in their formation, and therefore the subsequent heating or reduction of cooling may start from a state when there is practically no pressure or tension at the surface. Intrusion is therefore to be expected in mountainous regions. It is in agreement with this theory that large intrusions are habitually found there; they rarely reach the surface, but often come up to such a distance from the surface as to be accessible to the geologist. Such intrusions are of great horizontal extent and granitic in character; they are called *bathyliths*.

Other intrusions reach the surface in thin dykes and sills. They are not necessarily, nor indeed usually, associated with mountain ranges: they are usually basic. The acid character of bathyliths and the basic character of most (far from all) dykes may be due to the fact that an extra thickness of granitic matter is accumulated in a given region when a mountain range is formed; thus granite occurs below mountains at a depth where the matter below plains is basaltic.

A further type of elevation of fused matter from a considerable depth is afforded by volcanoes. These sometimes occur among folded mountains, as in the Andes, but more often in festoons of islands, especially in the Pacific. Their tendency to occur in chains indicates that they are formed along lines of weakness. It is clear, however, that though the volcanoes in a chain are related to the same line of weakness, they are not necessarily in communication with the same source of magma. If they were, lava in neighbouring volcanoes would rise to the same level, which is not the case: there is, for instance, a difference of 500 metres between the levels of the lava in two craters of Kilauea a few kilometres apart.

Tension phenomena in the earth's crust, though important, do not appear to be so general in character as the phenomena of folding. Bathyliths are generally considered by geologists to be subsidiary to mountain chains, and not conversely. A dyke is considered exceptionally long if it exceeds 100 miles. The importance of purely local considerations in determining the behaviour of volcanoes is shown by the difference in level between the two craters of Kilauea just mentioned; by the frequent changes of direction among the festoons of islands in the Pacific; and by the fact that as a rule different volcanoes of a chain are not active at one time. On the other hand the great mountain ranges, produced by folding, extend halfway round the earth. Thus tension phenomena must be regarded as due to local causes and compressional movements to worldwide ones.

Internal expansion, however, may produce bodily uplift of the overlying matter, without vertical fracture. This possibility, like the theory of intrusion just outlined, has not yet been submitted to quantitative test; but it appears to be indicated by some elevated plateaus with apparently little or no folding within them. Such regions are Tibet and the Deccan. They are of such dimensions, both vertical and horizontal, that our theory of the strength of the earth's crust requires them to be compensated. Below them, there must therefore be an abnormal thickness of light material. This could be produced by horizontal compression in two ways, both consistent with absence of folding in the uppermost layers; but neither is intrinsically very plausible. Firstly, the compression might have merely shortened and thickened the surface layer without crumpling it. This appears unplausible when we remember that the horizontal extent of these regions is of the order of 300 km., while the thickness of the granitic layer is of the order of only 15 km., so that crumpling is much more likely. The other possibility is that sheets of light matter from the sides were forced in below the undistorted matter already there. That this could be done without crumpling, in spite of the horizontal force to be overcome by the ends of the injected masses, appears quite as improbable as the first alternative. If, however, the matter below such a region, at depths of considerable strength, had spontaneously expanded, compensation would be attained without folding if the overlying rocks were merely forced up until the space between them and an equipotential surface within the region of zero strength was enough to accommodate the matter in its new state.

Some geodesists, notably Sir S. G. Burrard, have said that compensation is not produced or maintained by horizontal movement. In the account of Chapter IX it was shown that the establishment and maintenance of compensation by horizontal movement within the asthenosphere is a necessary inference from the theory of cosmogony that has been adopted, and that the main characteristics of the compensation inferred are in close accordance with those found by geodesy. Hence the geophysical theory here elaborated receives strong support from the facts of compensation. If, however, it were shown that compensation could be explained equally well in terms of another theory, it would not afford so strong an argument in favour of the present one The examination at this point of the possibility of compensation without horizontal movement is therefore desirable.

In the first place, the existence of folded mountains at all seems to demand horizontal compressive movements near the surface. Thus the surface motion alone would indicate an inflow of matter towards the region of mountain formation, and therefore an increase of mass within a vertical column. This can be compensated only if matter flows out from this column at greater depths: but this can be achieved only by horizontal

movement. Thus the compensation of mountains involves horizontal movement in the asthenosphere.

Next, consider the case of a plateau, originally compensated, in which a deep and broad valley has been cut by denudation. This involves a loss of mass in a vertical column, unless matter flows in horizontally below. Thus again horizontal movement within the asthenosphere is necessary. Similarly, the maintenance of the compensation of a river delta, as its area and thickness gradually increase, requires horizontal movement.

Burrard*, in reply to a criticism by Dr Morley Davies, has suggested that the compensation of a delta requires no depression of the crust. The matter below the crust merely acquires a lower density, without change of volume, and therefore the position of no element changes. This hypothesis, however, evidently contradicts the fundamental physical law of the indestructibility of matter.

Thus horizontal movement must play an essential part in any useful theory of mountain formation and of isostatic compensation. It is not asserted here that it has played the most important part in the formation of every surface feature; it is indeed suggested that certain plateaus require some other mechanism: but these are exceptional features, and horizontal compression accounts for many more and larger features than these.

* *Geographical Journal*, July 1920, p. 58.

CHAPTER XII

Seismology

"I heard the water lapping on the crag
And the long ripple washing in the reeds."

TENNYSON, *The Passing of Arthur.*

12·1. *Waves started by Fractures.* It has been seen that the earth's crust must be continually undergoing permanent deformation under the influence of the stresses developed within it in the course of its evolution. The deformation may take the form of gradual flow or of fracture. In each case strain energy is converted during the set into energy of internal motion. In the case of flow, the transformation is gradual, and produces only local heating; however great its ultimate effect may be, it produces no immediately observable consequences. The behaviour of the crust in the event of fracture, on the other hand, may be illustrated by analogy with an elastic string stretched to such an extent that it ultimately snaps. The relief of tension at the point of rupture initiates a separation of the exposed ends, and the motion spreads through the whole length of the string. In the same way, a fracture in the crust of the earth starts a movement there, leading to a wave, which spreads throughout the earth.

12·2. *Elastic Waves within a Solid: Longitudinal Waves.* It is easily seen that two types of wave are possible in a homogeneous elastic body. If u, v, w be the components of displacement from the equilibrium state, the equations of motion are

$$\rho \frac{\partial^2}{\partial t^2}(u, v, w) = (\lambda + \mu)\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\delta + \mu\nabla^2 (u, v, w) \quad \ldots\ldots(1),$$

where ρ is the density, λ and μ the two elastic constants of the substance, and

$$\delta = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \quad \ldots\ldots\ldots(2).$$

If a harmonic plane wave is being propagated with velocity c, the displacements must be the real parts of expressions of the form

$$(A, B, C)\, e^{\iota\kappa(x-ct)},$$

where the axis of x has been taken parallel to the direction of propagation, and A, B, and C are complex constants. By differentiating the three equations (1) with regard to x, y, z and adding, we find

$$-\rho\kappa^2 c^2 \delta = (\lambda + 2\mu)\nabla^2\delta \quad \ldots\ldots\ldots(3),$$

leading at once to

$$c^2 = (\lambda + 2\mu)/\rho \quad \ldots\ldots\ldots(4),$$

or

$$\delta = 0 \quad \ldots\ldots\ldots(5).$$

If δ is different from zero, then since $\partial\delta/\partial y$ and $\partial\delta/\partial z$ are zero, the second and third equations of motion give

$$\rho\kappa^2 c^2 (v, w) = \mu\kappa^2 (v, w) \qquad \text{.........(6)},$$

and since $\mu \neq \rho c^2$, v and w must be zero.

Thus there is a type of wave in which the displacement is wholly parallel to the direction of propagation, and the density of a particular element changes during the passage of the wave. This type of wave is known as the irrotational, longitudinal, condensational, primary, or 'P' wave. Its velocity c_1 is $\{(\lambda + 2\mu)/\rho\}^{\frac{1}{2}}$.

12·21. *Transverse Waves.* If, however, δ is zero, we have

$$\rho\kappa^2 c^2 (u, v, w) = \mu\kappa^2 (u, v, w) \qquad \text{.........(1)},$$

and therefore a necessary condition for a wave without condensation is that

$$c^2 = \mu/\rho \qquad \text{.........(2)}.$$

The wave must also satisfy the condition that δ, as defined by 12·2 (2), actually does vanish. Now

$$\frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \qquad \text{.........(3)},$$

in consequence of the assumption that the wave is plane. Hence

$$\frac{\partial u}{\partial x} = 0,$$

and therefore A is zero and there is no displacement parallel to the direction of propagation. Thus there is a second possible type of wave, such that the displacement is wholly perpendicular to the direction of propagation. There is no change in the density of any element during the passage of the wave. The displacements in two perpendicular directions parallel to the wave front are propagated independently. The difference of phase between the displacements parallel to these two directions is the same at all points, since both are propagated with the same velocity. Hence any particle may oscillate in a straight line, or may describe an ellipse or a circle, according to the amplitudes and phase differences of the components; but the size, form, and orientation of the path are the same for every particle. This type of wave is known as the distortional, transverse, equivoluminal, secondary, or 'S' wave. Its velocity c_2 is $(\mu/\rho)^{\frac{1}{2}}$.

In the case of secondary waves arriving at the surface, it is often desirable to distinguish between those whose displacements are horizontal and those whose displacements are in the plane including the vertical and the direction of propagation. It will be seen that the direction of propagation may be at any angle to the vertical. These two types will be called SH and SV waves. They are reflected in different ways, as will be seen from a paper by C. G. Knott*.

* *Phil. Mag.* Ser. 5, **48**, 1899, 64–97.

It can be proved that any motion whatever of an elastic solid under no external forces can be represented as the result of combining waves of these types.

12·3. *Rayleigh Waves.* In some possible movements of an elastic solid, the motion is greatest near the surface, and becomes inappreciable at a depth of a few wave lengths. Such waves can be represented as combinations of plane waves of condensational and distortional types rising from the interior, with the corresponding waves reflected downwards from the surface; but they are of such interest that a special discussion is desirable.

If the origin be in the surface, the axis of z vertically downwards, that of x horizontal and in the direction of propagation, and that of y perpendicular to it, then in any harmonic wave propagated over the surface without change of type the component displacements must be the real parts of $(U, V, W)\,e^{\iota\kappa(x-ct)}$, where U, V, W are complex quantities which may be functions of z. Then as in 12·2 (3) we find

$$\begin{aligned} -\rho\kappa^2 c^2 \delta &= (\lambda + 2\mu)\nabla^2\delta \\ &= (\lambda + 2\mu)\left(\frac{\partial^2\delta}{\partial z^2} - \kappa^2\delta\right) \qquad \text{.........(1).} \end{aligned}$$

The only solution of this equation that does not become infinite when z tends to infinity is

$$\delta = De^{-rz}\,e^{\iota\kappa(x-ct)} \qquad \text{.........(2),}$$

where D is a constant, and

$$r^2 = \kappa^2\left(1 - \frac{\rho c^2}{\lambda + 2\mu}\right) \qquad \text{.........(3).}$$

Any displacement consistent with this value of δ must satisfy

$$-\rho\kappa^2 c^2(U, V, W) = (\lambda + \mu)(\iota\kappa, 0, -r)De^{-rz} + \mu\left(\frac{d^2}{dz^2} - \kappa^2\right)(U, V, W) \quad \text{...(4),}$$

since when R is of the form $Pe^{\iota\kappa(x-ct)}$, where P is a function of z alone,

$$\nabla^2 R = \left(\frac{d^2}{dz^2} - \kappa^2\right)Pe^{\iota\kappa(x-ct)} \qquad \text{.........(5).}$$

Also $\dfrac{\partial u}{\partial x} + \dfrac{\partial v}{\partial y} + \dfrac{\partial w}{\partial z}$ is the real part of

$$\left(\iota\kappa U + \frac{dW}{dz}\right)e^{\iota\kappa(x-ct)} \qquad \text{.........(6),}$$

and must be equal to $De^{-rz}\,e^{\iota\kappa(x-ct)}$.

Particular solutions of these equations are found to be

$$\begin{aligned} (U, V, W) &= -\frac{\lambda + 2\mu}{\rho\kappa^2 c^2}(\iota\kappa, 0, -r)De^{-rz} \qquad \text{.........(7)} \\ &= (U_1, V_1, W_1),\ \text{say} \qquad \text{.........(8).} \end{aligned}$$

It is readily verified that

$$\iota\kappa U_1 + \frac{\partial W_1}{\partial z} = De^{-rz} \qquad \text{.........(9).}$$

Hence the complete solution, satisfying

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=\delta \qquad \text{.........(10)},$$

must be given by

$$(U, V, W)=(U_1, V_1, W_1)+(U_2, V_2, W_2) \qquad \text{.........(11)},$$

where (U_2, V_2, W_2) satisfy the equations

$$-\rho\kappa^2c^2(U_2, V_2, W_2)=\mu\left(\frac{d^2}{dz^2}-\kappa^2\right)(U_2, V_2, W_2) \qquad \text{......(12)},$$

and

$$\iota\kappa U_2+\frac{dW_2}{dz}=0 \qquad \text{.........(13)}.$$

The solutions of these equations that tend to zero at an infinite depth are

$$(U_2, V_2, W_2)=(s, \beta, \iota\kappa)\,Qe^{-sz} \qquad \text{.........(14)},$$

where Q and β are unspecified constants and

$$s^2=\kappa^2\left(1-\frac{\rho c^2}{\mu}\right).$$

Now the boundary conditions are that there shall be no stress across the free surface; and if we neglect small quantities of the second order in the displacements this is equivalent to the condition that there shall be no stress over the plane $z=0$. This gives

$$\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}=0 \qquad \text{.........(15)},$$

$$\frac{\partial v}{\partial z}+\frac{\partial w}{\partial y}=0 \qquad \text{.........(16)},$$

$$\lambda\delta+2\mu\frac{\partial w}{\partial z}=0 \qquad \text{.........(17)},$$

when $z=0$. Substituting in these equations from (8) and (14), we have

$$2\iota\kappa r\,\frac{\lambda+2\mu}{\rho\kappa^2c^2}\,D=(\kappa^2+s^2)\,Q \qquad \text{.........(18)},$$

$$\beta=0 \qquad \text{.........(19)},$$

$$(\lambda+2\mu)\left(1-\frac{2\mu}{\rho c^2}\right)D=2\mu s\iota\kappa Q \qquad \text{.........(20)}.$$

Eliminating D and Q, we have as an equation to find c,

$$\left(1-\frac{2\mu}{\rho c^2}\right)(\kappa^2+s^2)+4\mu\frac{rs}{\rho c^2}=0 \qquad \text{.........(21)},$$

or, eliminating κ, and writing

$$\frac{\lambda+2\mu}{\rho}=c_1{}^2, \quad \frac{\mu}{\rho}=c_2{}^2, \qquad \text{.........(22, 23)},$$

$$\left(2-\frac{c^2}{c_2{}^2}\right)^2=4\left(1-\frac{c^2}{c_1{}^2}\right)^{\frac{1}{2}}\left(1-\frac{c^2}{c_2{}^2}\right)^{\frac{1}{2}} \qquad \text{.........(24)},$$

or, on squaring, $$\left(2-\frac{c^2}{c_2^2}\right)^4 = 16\left(1-\frac{c^2}{c_1^2}\right)\left(1-\frac{c^2}{c_2^2}\right) \qquad \text{.........(25),}$$

which reduces to $$\frac{c^6}{c_2^6} - 8\frac{c^4}{c_2^4} + 24\frac{c^2}{c_2^2} - 16\frac{c^2}{c_1^2} - 16 + 16\frac{c_2^2}{c_1^2} = 0 \text{ ....(26).}$$

When the solid is incompressible, c_1 is infinite, and the equation reduces to

$$\frac{c^6}{c_2^6} - 8\frac{c^4}{c_2^4} + 24\frac{c^2}{c_2^2} - 16 = 0 \qquad \text{.........(27).}$$

This has one real root, $$c^2 = 0{\cdot}91275c_2^2 \qquad \text{.........(28).}$$

The two complex roots make

$$c^2/c_2^2 = 3{\cdot}5436 \pm 2{\cdot}2301\iota \qquad \text{.........(29),}$$

and then (24) gives $$4s/k = -\,2{\cdot}7431 \pm 6{\cdot}8846\iota \qquad \text{.........(30),}$$

so that e^{-sz} tends to infinity with depth. Thus these roots are inadmissible. Hence the only type of wave possible is determined by

$$c = 0{\cdot}9554c_2; \quad s = 0{\cdot}2954\kappa; \quad r = \kappa \qquad \text{.........(31),}$$

$$u = A\,(e^{-\kappa z} - 0{\cdot}5433e^{-sz}) \sin\kappa\,(x-ct) \qquad \text{.........(32),}$$

$$v = 0 \qquad \text{.........(33),}$$

$$w = A\,(e^{-\kappa z} - 1{\cdot}840e^{-sz}) \cos\kappa\,(x-ct) \qquad \text{.........(34),}$$

where A is a constant. The velocity of propagation is rather less than that of secondary waves. There is no displacement across the direction of propagation. The motion at the surface is given by

$$u = 0{\cdot}4567A \sin\kappa\,(x-ct) \qquad \text{.........(35),}$$

$$w = -\,0{\cdot}840A \cos\kappa\,(x-ct) \qquad \text{.........(36),}$$

so that the particles move in elliptic paths, the maximum vertical and horizontal displacements being in the ratio 1·9 : 1.

If Poisson's ratio is $\frac{1}{4}$, so that $\lambda = \mu$, Lord Rayleigh finds

$$c = 0{\cdot}9194c_2 \qquad \text{.........(37),}$$

$$s = 0{\cdot}3933\kappa \qquad \text{.........(38),}$$

$$r = 0{\cdot}8475\kappa \qquad \text{.........(39),}$$

$$u = A\,(e^{-rz} - 0{\cdot}5773e^{-sz}) \sin\kappa\,(x-ct) \qquad \text{.........(40),}$$

$$w = A\,(0{\cdot}8475e^{-rz} - 1{\cdot}4679e^{-sz}) \cos\kappa\,(x-ct) \text{ ....(41),}$$

and the ratio of the axes of the elliptic orbit described by a surface particle is reduced to about $\frac{3}{2}$.

The type of wave just discussed was discovered by Lord Rayleigh, and is usually named after him. It has been seen in the course of the investigation that no other type of harmonic wave is capable of continuous propagation over the surface of a homogeneous solid; and that in this type the displacement of any particle is partly vertical and partly horizontal in the direction of propagation, there being no horizontal motion across the direction of propagation.

12·4. The actual earth being heterogeneous, its vibrations must be more complicated than those examined in paragraphs 12·2 to 12·3; a complete theory of them has not yet been constructed. Certain essential points, however, can be made clear without elaborate analysis. All the types of wave so far considered travel with velocities independent of the wave length. Now every wave train, of whatever form, can in general be expressed as the resultant of a number (usually, in a continuous body, an infinite number) of simple harmonic wave trains. Again, whatever the nature of the initial disturbance that generates the disturbance, it can be expressed as the resultant of a number of simple harmonic motions. If then it takes T seconds for one of these waves to travel from the origin to a given particle, it will follow that at any time the motion at the particle will be in the same phase as at the origin T seconds before. Further, if the waves all travel with the same velocity, this will be true for every wave separately. Hence, in accordance with the theory of group velocity, the motion of the particle will reproduce the motion at the origin T seconds previously, but reduced in amplitude on account of the fact that each wave is diverging all round. Thus a single impulse at the origin will produce a single impulse at any other point after a definite interval depending on the wave velocity and the distance. In fact, the waves would pass a point of the surface in three stages: first, the compressional wave would pass, then the distortional wave, and then the Rayleigh wave, each producing a sudden jerk, without much motion at any other time.

12·41. In the heterogeneous earth the impulse at the origin will generate primary and secondary waves in its neighbourhood, and these will spread out with the velocities appropriate to their types. Every time they enter matter with different physical properties, however, they will undergo modification. In part they will be reflected, and in part they will be transmitted, with change of velocity and direction of propagation. Each point of the boundary between the two media will act as a new origin, and the wave will first affect a given point in the second medium after a time equal to the shortest possible time of transmission of a wave of the given type from the origin to the point. The principle of least time being the basis of the ordinary laws of refraction, we see that the wave that initiates each phase of the disturbance observed at any station must have traversed from the origin the path specified by the laws of refraction. If we proceed to the case of several layers, and finally to that of an earth whose properties vary continuously from point to point, the same results will evidently hold. Thus an impulse applied within the earth will generate two main waves of primary and secondary type, spreading out with the velocities appropriate to waves of these types, and sudden displacements of any particle will occur when these waves reach it. Since, however, the velocities are not the same in all parts of the earth, the wave front at any instant will not be spherical. Internal reflexion will produce waves which

can reach the point of observation only by more circuitous routes than the two main waves, and thus such reflected waves will arrive later. Where there is a definite boundary between different materials, definite reflected waves will be produced, which will be capable of being recorded separately at the observing station; but where the transition is continuous, these pulses will be distributed over a long interval, and will not give definite sudden shocks.

12·42. It is actually found that the first effect of an earthquake at a distant station is a sudden impulse in the direction away from the place where the earthquake has actually occurred. This corresponds to the primary wave, just discussed. This is followed by a further sudden impulse, corresponding to the arrival of the secondary wave. Before and after the latter come trails of waves that can be identified with waves reflected internally, some having been changed from longitudinal to transverse by internal reflexion, and some from transverse to longitudinal. Thus the actual phenomena are in close accordance with what would theoretically be expected. The times of arrival of the P and S waves have constituted the greater part of the data of seismology; less attention has been given to the trains of internally reflected waves, partly on account of the difficulty of describing them without reproducing the actual records of the earthquakes, which would be an extremely expensive operation to undertake if more than a few earthquakes were to be discussed.

12·43. The waves will undergo reflexion, not only at internal points, but also when they reach the surface. The theory of the reflexion of elastic waves at the surface of a solid is somewhat complex, but has been developed by Knott in the paper already cited. In general a pure condensational wave or a pure distortional wave generates waves of both types on reflexion, the amplitudes of the condensational and distortional reflected waves depending on whether the incident wave was condensational or distortional, on the plane of polarization of the incident wave, where this is distortional, and on the angle of incidence. Knott's numerical results for the reflexion of waves from an interface between rock and air are as follows. The density of the rock has been taken to be 2000 times that of the air.

(1) *Distortional wave incident in the rock.*

ϕ	B	B_1	θ	A_1	θ'	A'
0	1	1	0		0	
14° 2′	1	0·534	25°	0·466	1°·6	0·00002
26° 34′	1	0 025	51°	0·975	3°	0·00006
33° 40′	1	0·003	74°	0·997	3°·7	0·00006
35° 13′	1	1	90°	0 000	3°·8	0·00000
39° 48′	1	1		—	4°·3	0·00019
45° 0′	1	1		—	4°·7	0·00016
59° 2′	1	1		—	5°·7	0·00014
73° 18′	1	1		—	6°·3	0·00014
84° 17′	1	1		—	6°·6	0 00006

(2) *Condensational wave incident in the rock.*

θ	A	A_1	θ'	A_1'	ϕ'	B_1
0	1	1	0	0·00013	0	0
14° 2′	1	0·828	0°·9	0·00013	8°	0·172
26° 34′	1	0·464	1°·7	0·00011	15°	0·536
45° 0′	1	0·079	2°·7	0·00009	24°	0·921
59° 2′	1	0·002	3°·3	0·00007	30°	1·000
73° 18′	1	0 003	3°·7	0·00006	34°	0·997
84° 17′	1	0·091	3°·8	0·00005	35°	0·909

In these two tables the symbols without accent or suffix refer to the incident wave, those with suffix 1 to the reflected wave in the solid, and those with accents to the wave refracted into the air. θ is the angle of incidence, refraction, or reflexion of a wave of condensational type, and A its energy; ϕ is the angle of incidence or reflexion of a wave of distortional type, and B its energy. The absence of a distortional wave in air ensures that B' is always zero.

It is seen that in no case does more than a small fraction of the energy of the incident wave pass out into the air. Thus the sound-wave produced by an earthquake must in all cases carry an amount of energy very small in comparison with that of the earthquake itself.

In calculating Table (1), Knott has supposed the vibration of the incident wave to be in the vertical plane through the direction of propagation; that is, it is a wave of the type already denoted by SV. This produces a vertical motion of the boundary, and therefore can give rise to a wave in the overlying fluid. It is seen that as a rule most of the energy of such a wave is reflected in a distortional wave, except for angles of incidence between 15° and 30°. In this range the reflected wave is mostly condensational, the transition to the condition of reflexion as a pure distortional wave being extraordinarily sudden. Thus a wave of type SV incident normally, or at a direction of propagation making an angle with the normal greater than 35°, is reflected as a pure wave of type SV, whereas if its angle of incidence is within the range from 15° to 30°, the reflected wave is mostly of type P.

A wave of type SH gives no vertical movement of the boundary, and therefore no wave in the air. The stress and strain in such a wave are purely horizontal, and it is therefore evident that the condition that the stress shall be continuous across the surface can be satisfied by combining with the incident wave a reflected wave of the same amplitude and with an angle of reflexion equal to the angle of incidence, the phase being such as to neutralize the surface stress that would be produced by the incident wave acting alone. Thus waves of type SH are reflected as pure SH waves, without giving any condensational waves.

Table (2) shows that when a condensational wave is reflected, most of the energy is reflected in the condensational wave for directions of propagation making angles less than 20° with the vertical, but that for

greater angles of incidence the reflected wave of type SV carries most of the energy.

Waves reflected at the surface differ from those reflected in the interior, since we definitely know that they are reflected at a given surface, whereas waves reflected internally may be, and probably usually are, reflected continuously through regions of finite thickness, within which the physical properties of the medium change continuously. The continuous character of internal reflexion renders it unlikely that it would give reflected waves recognizable by definite impulses on the seismic records. In the case of surface reflexion, on the other hand, perfectly definite waves of the same type as the original waves are produced by reflexion, and must afterwards be propagated according to the same laws. When they again emerge at the surface they therefore give definite impulses which are recognizable on the records. In general the impulse due to a wave that started as a condensational one, and was reflected as a condensational one, is denoted by PR; that due to a wave originally condensational, but reflected as distortional, by PS; that due to the condensational wave produced by the reflexion of a distortional wave by SP; and that due to a distortional wave reflected as distortional by SR. In addition each type of reflected wave may undergo further reflexion when it again emerges. The waves produced by these later reflexions are more difficult to identify on the records, but it is probable that a phenomenon known as the Y wave is due to them.

12·5. *The Long Waves.* The seismograph records a sudden movement of the ground when any of these well-defined waves first reaches it. Between them the instrument is disturbed only by waves reflected internally. The sudden movement that initiates a new phase marks the instant of arrival of a wave of given type after the shortest possible time of transit. The internally reflected waves, however, will also undergo reflexion when they emerge at the surface. In general the reflected waves formed in this way will be even more difficult to identify than the internally reflected waves are when they emerge for the first time, and therefore will afford no prospect of successful discussion in the present state of seismological knowledge. A portion of the wave will be diffusely reflected, since the solid surface of the earth is irregular in contour and in constitution. Thus, wherever the origin of the earthquake may be, we should expect that some part of its energy will be reflected at the surface and sent out along the surface. It will not, in general, be capable of proceeding far along the surface without loss into the interior. It has been seen, however, that Rayleigh waves, if once started, can proceed over the surface for an indefinite distance without change of type or dispersion into the interior. Thus we may expect that part of the wave arriving at the surface will be diffusely reflected, and that some of the motion will go to produce further vibration in the

interior; but part of the energy of the wave will be used up in the formation of surface waves. Now a given amount of energy, in spreading through the earth without reflexion, becomes distributed over an area roughly proportional to the square of the distance from the source, so that the energy crossing a given area is proportional to the inverse square of the distance. On the other hand, in a surface wave propagated without change of type, the distribution with regard to depth remains unaltered as the wave advances, and the area over which the energy is spread is roughly proportional to the circumference of a small circle of the earth, with its centre above the origin of the earthquake, and passing through the point of observation. Thus the energy crossing a given small area near the surface is proportional only to the inverse first power of the radius of this circle. Hence the importance of the surface waves, in comparison with that of the internal waves, will gradually increase with distance from the origin, and it is possible that for stations sufficiently remote they will be responsible for most of the observable motion.

Now it is observed that at stations far away from the origin of the earthquake the arrival of the S wave is followed by a stage in which the motion is oscillatory in character, the amplitude and period varying only slowly. The amplitude increases to a maximum usually several times greater than the greatest displacement in the P and S stages, and then falls again. Often, however, the amplitude rises and falls several times before quiescence is again attained. This stage is known as the long wave or L phase, or as the main shock. Since Rayleigh's investigation showed that a surface wave was capable of producing at remote stations a disturbance greater than either P or S, this stage of an earthquake has been generally interpreted as the passage of Rayleigh waves. Such an interpretation, however, can account for only part of the facts. In the first place, it has been seen that the velocity of Rayleigh waves in a homogeneous earth is independent of the wave length, and is a fixed proper fraction of the velocity of S. Hence, whatever periods may occur in the waves reflected from the surface, all should reach any given station at the same time, and the Rayleigh phase should consist of one shock, and one only; the simple theory cannot account for a long train of approximately harmonic waves. Again, the motion in the Rayleigh wave is wholly in the plane containing the vertical and the direction of propagation. The actual motion in the L phase has a strong horizontal component at right angles to the direction of propagation.

12·51. Both difficulties have been largely removed by the investigation of Love* on the propagation of elastic waves in a uniform layer of finite depth, resting on another uniform layer of infinite depth. He has shown, first, that the behaviour of Rayleigh waves is decidedly different from that

* *Problems of Geodynamics*, Camb. Univ. Press, 1911.

in a uniform crust. The velocity, instead of being constant, depends on the period. The truth of this statement is qualitatively evident without analysis, for a wave of short period will have a short wave length, and therefore, in view of the fact that r and s in 12·3 were seen to be proportional to κ, a short vertical extent. Thus if the period is sufficiently short the motion will be practically confined to the upper layer, and the velocity of Rayleigh waves will be that characteristic of that layer. On the other hand, if the period is long enough the wave will extend so far into the lower layer that the small thickness of the upper layer will not affect the motion appreciably, and the velocity of Rayleigh waves will be that characteristic of the lower layer. Hence the velocity of Rayleigh waves in a heterogeneous crust will depend on the period. The initial disturbance being capable of representation as a combination of waves with a wide range of periods, it is therefore clear that the waves of different periods will spread out at different rates and pass a given station at different times. The motion at a distant station will therefore not be of the nature of a sudden jerk, but will consist of a long series of oscillations of gradually varying period. This is just what is observed.

12·52. *Love Waves.* Love finds that even in a heterogeneous crust Rayleigh waves have no transverse horizontal component. On the other hand, he has shown the possibility of a type of wave impossible in a homogeneous crust, in which the displacement is purely horizontal and at right angles to the direction of propagation.

Suppose the lower boundary of the surface layer to be the plane $z = 0$, and the upper to be the plane $z = -T$. The wave is being propagated parallel to the axis of x. Let μ and ρ be the rigidity and density of the upper layer, and let μ' and ρ' refer to the lower layer. Then in both layers u and w are zero, while v is proportional to the real part of $Ve^{\iota\kappa(x-ct)}$, where V may be complex, but is a function of z alone. In both layers δ is zero, and the only equation of motion that is not satisfied identically is

$$-\rho\kappa^2c^2V = \mu\left(-\kappa^2V + \frac{\partial^2 V}{\partial z^2}\right) \qquad \text{.........(1)}$$

for the upper layer, and

$$-\rho'\kappa^2c^2V = \mu'\left(-\kappa^2V + \frac{\partial^2 V}{\partial z^2}\right) \qquad \text{.........(2)}$$

for the lower layer. Putting, as before, $\mu/\rho = c_2^2$, $\mu'/\rho' = c_2'^2$, we have in the lower layer

$$V = Ce^{-s'z} \qquad \text{.........(3),}$$

where C is a constant and

$$\frac{s'^2}{\kappa^2} = 1 - \frac{c^2}{c_2'^2} \qquad \text{.........(4).}$$

In the upper layer

$$V = A\cos sz + B\sin sz \qquad \text{.........(5),}$$

where A and B are constants and

$$\frac{s^2}{\kappa^2} = \frac{c^2}{c_2^2} - 1 \qquad \text{.........(6).}$$

The boundary conditions are that the displacement and the tangential stress are continuous across the lower boundary, and that there is no stress across the free surface. These give

$$C = A \qquad \text{.........(7)},$$

$$-\mu' s' C = \mu s B \qquad \text{.........(8)},$$

$$A \sin sT + B \cos sT = 0 \qquad \text{.........(9)}.$$

Hence by elimination of A, B, C we have

$$\tan sT = \mu' s'/\mu s \qquad \text{.........(10)}.$$

Finally, by substituting for s and s' from (4) and (6), we have the following equation for the wave velocity,

$$\tan \kappa T \left(\frac{c^2}{c_2^2} - 1\right)^{\frac{1}{2}} = \frac{\mu'}{\mu}\left(\frac{1 - c^2/c_2'^2}{c^2/c_2^2 - 1}\right)^{\frac{1}{2}} \qquad \text{.........(11)}.$$

The condition that the motion shall be indefinitely small at a great depth requires that the real part of s' shall be positive. Hence, by (4), c must be less than c_2'. Again, if c were less than c_2, s would be a pure imaginary, by (6), and therefore $s \tan sT$ would be negative. This, by (10), would imply that s' was negative, which is impossible. It follows that c is always greater than c_2 and less than c_2'. There can therefore be no waves of this type if c_2 is greater than c_2', for it would then be impossible to find a value of c that would satisfy both of these inequalities. Thus these waves can exist only if the velocity of distortional waves in the surface layer is less than in the matter below.

An optical analogy will illustrate the reason for this result. A wave moving in a surface layer may be compared to a light wave moving nearly parallel to the faces of a plate, with matter of a different refractive index in optical contact with the lower face and a perfect reflecting surface over the upper face. If the velocity of light in the plate is less than that in the underlying matter, light striking the boundary at angles less than the critical angle will be totally reflected, and will continue to be propagated in the plate. If, however, the velocity of light is greater in the plate, light can pass freely from the plate to the underlying matter, whatever its angle of incidence may be. Thus the light will gradually be lost into the underlying material as it advances, and the propagation of a train of waves of constant form will be impossible.

Equation (10) can be written

$$\sigma \tan \sigma\kappa T = \mu'\sigma'/\mu \qquad \text{.........(12)},$$

where σ and σ' have been written for $(c^2/c_2^2 - 1)^{\frac{1}{2}}$ and $(1 - c^2/c_2'^2)^{\frac{1}{2}}$ respectively. When c is equal to c_2, σ is zero, while σ' is finite and positive. Thus whatever κT may be the left side of (12) will be zero and the right positive. As c increases, the left side will increase steadily until $\sigma\kappa T$ becomes equal to $\frac{1}{2}\pi$, when the left side becomes infinite. If then $(c_2'^2/c_2^2 - 1)^{\frac{1}{2}}\kappa T$ is greater than $\frac{1}{2}\pi$, the left side will become infinite and

positive for some value of c between c_2 and c_2'. It will then become negative and again increase steadily; if $(c_2'^2/c_2^2 - 1)^{\frac{1}{2}} \kappa T$ is greater than π but less than $\frac{3}{2}\pi$, the left side will again be positive when c is equal to c_2'; the discussion can evidently be extended indefinitely. The right side meanwhile decreases steadily from $\mu' (1 - c^2/c_2'^2)^{\frac{1}{2}}/\mu$ to zero. Thus the left side will exceed the right when c is equal to c_2', unless it has become negative at some intermediate point by passing through an infinite value and has not become positive again by passing through zero; in either case there will be some intermediate value where the left side exceeds the right, and therefore some other value of c that makes the two sides equal. The condition for this is that $(c_2'^2/c_2^2 - 1)^{\frac{1}{2}} \kappa T$ shall be less than π. Similarly we see that there will be two possible values of c that satisfy (12) if $(c_2'^2/c_2^2 - 1)^{\frac{1}{2}} \kappa T$ lies between π and 2π, three if this quantity lies between 2π and 3π, and so on. Again, we see readily that increasing κ while keeping T, μ and μ' the same makes all the roots approach c_2. Thus the shorter the wave length, the more closely will the velocity approach that of distortional waves in the upper layer; and when it becomes indefinitely short these velocities tend to equality. Again, if κ is indefinitely small, we see that the equation (12) can be satisfied only when c approximates to c_2'; thus waves of great length will travel with velocities approximating to that of distortional waves in the lower layer.

When more than one value of c corresponds to the same value of κ, sz changes by more than π as z changes from 0 to $-T$; hence V, being expressed as a harmonic function of sz with real coefficients, must vanish for some intermediate value of z. Thus there will be no, one, two, or more nodal surfaces within the upper layer according as the root considered is the lowest, second, third or higher value of c corresponding to the actual wave length.

We have seen that A, B, and C are all real; it follows that the motion is in the same phase at all depths.

12·53. The type of waves just discussed will be referred to as Love waves. It has been seen that the displacement in them is wholly horizontal and at right angles to the direction of propagation, they should not affect the vertical component. This is in agreement with the appearance of the records. A seismogram of the vertical movement shows a smooth oscillation, which is readily attributable to a single type of wave, namely the Rayleigh waves; but the two horizontal components show violent irregularities, indicating the passage of two sets of waves of widely different periods. It seems probable that the slower of the oscillations shown by the horizontal components are due to Rayleigh waves and to those Love waves that have no nodal plane, while the rapid ones are due to the Love waves with nodal planes. Since all the waves passing one place at one instant must have started from the origin at the same instant, their

velocities must all be equal. Thus in equation (12) all the quantities except κ must be the same for all, and all the values of κ must differ by multiples of $\pi/\sigma T$.

The analysis of the long wave phase from the observational data has hitherto received very little attention. The general agreement of its character with that inferred from Love's and Rayleigh's analyses is enough to indicate that the theory is substantially correct, but the observational results could certainly be made to yield much more information about the nature of the upper regions of the earth's crust than they have yet provided. For instance, the analysis of the periods of the waves arriving at a given time should, by the rule given in the last paragraph, lead to an estimate of the thickness of the superficial layer.

12·54. The foregoing account of the theory of the propagation of earthquake waves has been developed directly from the theory of elastic waves in a solid, the observational data having been used only for purposes of qualitative comparison with the results of the theory. The main inferences that have so far been found necessary to harmonize theory and observation are that there must be a level at a moderate depth in the earth's crust where the velocity of propagation of distortional waves undergoes a considerable increase, and that there must be enough heterogeneity within the earth to produce a considerable amount of internal reflexion. The former result harmonizes well with the result of the theory of the cooling of the earth, which suggested that the granitic layer of the continents was probably underlaid by basic rocks of much greater rigidity at a depth of the order of 10 to 20 km.

12·6. *Determination of the Velocities of Propagation of the* P *and* S *Waves.* We now come to consider the additional information supplied by a more detailed study of the data that have hitherto chiefly engaged the attention of seismologists, namely, the times of arrival of the P and S waves. Since the time of transit of a wave depends only on its path and on its velocity in the matter at each point of its path, it is clear that a knowledge of the times of transit of waves between many known points of the earth would furnish valuable information about the velocities of these waves along the paths they traverse. The actual solution of the problem is, however, one of considerable difficulty. The actual earthquake does not take place at a known place; it happens at an uncertain depth below the surface, and the time of the rupture is never directly observed, but has to be inferred from the times of the P and S shocks at the observing stations themselves. Thus the observations of every earthquake involve two unknowns in addition to the velocities of the waves within the earth, and these two unknowns are different for every earthquake. Accordingly the reduction of the observations for analysis is bound to be a matter of great difficulty.

12·61. There are two instances where the situation and depth of the origin are accurately known, though neither shock is an earthquake in the narrowest sense of the term. These are the Oppau explosion of 1921 Sept. 21, and the Pamir landslip of 1911 Feb. 18. The former was recorded on seismographs at distances up to 365 km. from the origin; the latter at distances from 400 km. up to 11,000 km. Together they therefore give data for the times of propagation of earthquakes over all distances up to over a quadrant, while the depth of the origin in each case is known to be accurately zero. Thus the chief source of uncertainty is absent from these two shocks.

The Oppau explosion* took place at the works of the Badische Anilin und Sodafabrik at Oppau, in the Bavarian Palatinate, on 1921 Sept. 21. Oppau is about 5 km. north-west of Mannheim, and stands on the alluvium of the Rhine valley. The shock was recorded at several seismographic stations, the results being as follows:

Station	Distance from Oppau (km.)	P h. m. s.	S h. m. s.
Strasbourg	110	6 32 33	6 32 50
Nordlingen	175	32 44	33 5
Zurich	240	32 58	33 28
Munchen	282	33 6–12	33 42
De Bilt	365	33 19	—

The times are all Greenwich mean time. In all cases except Nördlingen they refer to the mean of the components available. The Nördlingen observation refers only to the E.W. component. This is a smoked-paper record, but is of remarkable clearness; the motion is very small in extent, indeed almost invisible to the unaided eye, but under a magnifying glass all the phases are astonishingly distinct. Strasbourg obtained good records of all three components. The records at the more distant stations are naturally less in amplitude, but the times are fairly clear. The reason for the uncertainty in the time of P at München is that it came during the gap on the record that records the end of the minute, the time correction being + 6 seconds. De Bilt failed to record S.

The chord from Oppau to De Bilt would penetrate only about 3 km. below the surface of the earth, and the influence of the curvature of the earth on the results will therefore be unimportant. The time of the explosion is not accurately known; let us suppose it to be 6h. 32m. Ts., where T is to be found. Let the times taken by P and S waves to travel 100 km. be p and s respectively. Then the times of the observed shocks give the conditions

$$\begin{array}{ll} T + 1{\cdot}10p = 33, & T + 1{\cdot}10s = 50, \\ T + 1{\cdot}75p = 44, & T + 1{\cdot}75s = 65, \\ T + 2{\cdot}40p = 58, & T + 2{\cdot}40s = 88, \\ \cdots\cdots\cdots\cdots & T + 2{\cdot}82s = 102, \\ T + 3{\cdot}65p = 79. & \cdots\cdots\cdots\cdots \end{array}$$

* Dorothy Wrinch and Harold Jeffreys, *M.N.R.A.S. Geophys. Suppt.* 1, 1923, 15–22.

The solution of these equations by the method of least squares gives

$$T = 13 \text{ secs.};\quad p = 18{\cdot}5 \text{ secs.}/100 \text{ km.};\quad s = 31{\cdot}7 \text{ secs.}/100 \text{ km.}$$

The residuals (observed time − calculated time) are as follows, in seconds:

	Strasbourg	Nordlingen	Zurich	Munchen	De Bilt
P	0	− 1	1	1–7	− 1
S	2	− 3	− 1	0	—

The agreement of all the observations with the theory within a few seconds must be regarded as extremely satisfactory. Our hypothesis that they all refer to two waves spreading out with uniform velocities in a homogeneous medium is therefore in close accordance with the facts. The velocities indicated are 5·4 km./sec. for P and 3·1 km./sec. for S.

12·62. On 1911 Feb. 18 a huge mass of rock fell from a mountain in the Pamir region and transformed a river valley into a lake. A subsequent survey by Lieut. Spilko, of the Russian Army, formed the basis of a determination by M. Weber, of the Russian Geological Survey, of the size of the mass and the height it fell through. The fallen mass weighed 7×10^{15} to 10^{16} grams, indicating a volume of 2 or 3 cubic kilometres, and fell 300 to 600 metres. The coordinates of the spot where the catastrophe took place were 38° 16′ N., 72° 34′ E.

A large earthquake took place on this day, and was recorded at seismological stations all over the earth. Galitzin* found that the times of arrival of the P and S waves at Pulkovo were consistent with the origin having coincided with the landslip, and with a time of the earthquake within three minutes of that of the landslip, the time of the landslip being itself uncertain by quite this amount. Subsequent examination of the Ottawa records by O. Klotz†, and of the Eskdalemuir records by myself, showed that these also were consistent with these coincidences, so that there seems no room for doubt that the earthquake recorded by the seismographs agreed so closely in situation and time with the landslip that a dynamical connection was highly probable. It was possible, on the one hand, that the blow to the ground caused by the fall might have been the cause of the seismic disturbance that travelled out from the place, part of the kinetic energy acquired by the mass during its fall being converted into energy of internal vibration in the earth. On the other hand, it was possible that the earthquake originated at some depth, and that the rock mass was only loosened by the shock; if this were so, the blow to the ground would be only an incidental effect of a much larger disturbance. In the former case, the energy communicated to the earth by the impact should have been equal to that which spread out in the wave; while in the latter case the energy of the wave should much exceed that produced by the fall. Hence the evaluation of these two energies should enable us

* Prince B. Galitzin, *Comptes Rendus*, **160**, 1915, 810–13.

† O. Klotz, *Journal of the R.A.S. of Canada*, **9**, 1915, 428–37.

to decide between these two theories of the landslip. Galitzin, in the paper just quoted, attempted to determine these two energies, and obtained concordant results; he inferred that the wave was caused by the landslip, and not *vice versa*. Unfortunately he used incorrect theories in making both his estimates, but the evaluation has been revised by the present writer*.

Only a small fraction of the energy of the falling body would go into the wave; the greater part would be used in fracturing the body itself. The fraction available for the earthquake has been evaluated, and found to be of the order of $0{\cdot}8 \times 10^{21}$ to $2{\cdot}4 \times 10^{21}$ ergs. The energy actually in the earthquake appears to have been about 2×10^{21} ergs; Galitzin overestimated it through assuming that the long waves were equally intense throughout the interior, an assumption for which there is no theoretical justification. The agreement between the two estimates is as close as could be expected. It may be pointed out that this fact is sufficient to show that the energy of the wave was derived from the landslip; exact agreement is not necessary to the establishment of this point. For the argument is enough to show that the landslip must have produced a wave of magnitude comparable with the observed wave. If the landslip was only a consequence of a previous wave from a deep focus, the wave would also be recorded on the records at an earlier time. Hence the P and S shocks would be duplicated at intervals corresponding to the depth of the origin. This is not the case. Thus the hypothesis of a deep-seated focus becomes untenable when it is realized that the wave caused by the landslip must in any case have been large enough to be recorded separately.

Accordingly we have in the Pamir landslip an earthquake of the first magnitude whose origin was in the surface. Unfortunately, however, we have not the other datum that would be necessary to enable us to make the fullest use of the records, namely, the time of the shock. In the case of the Oppau explosion we could fix this with an error probably not exceeding a second, since the observing stations were all so near the origin that the waves in transit had been confined to a single layer of the crust. The nearest station that recorded the Pamir landslip was Tashkent, 440 km. away. This is indeed not much greater than the distance of De Bilt from Oppau, so that the results of the Oppau explosion, applied to the Tashkent observations, should enable us to fix the time of the fall, assuming that the crust in the Pamir region is made of similar materials to that in the Rheinland, and that the times observed at Tashkent are reliable within a few seconds. Both hypotheses are open to criticism, the latter especially, since the international wireless time service had not in 1911 begun to supply correct time with the accuracy it had reached in 1921. Accordingly, it would be unsafe to adopt an estimate of the time of the fall based on the Oppau velocities and the times of P and S observed

* *M.N.R.A.S. Geophys. Suppt.* 1, 1923, 22–31.

at Tashkent, and to infer from this and the observed times at other stations the times it takes the waves from a surface shock to travel distances up to 11,000 km. The Pamir times can be used only as a check on, and supplement to, other data, and are not in themselves sufficient to serve as the basis of quantitative seismology.

12·63. The other, and more usual, method of attack on the problem of the times of transmission of the P and S waves is by statistical treatment of the records of earthquakes. This method is open to the objection of 12·6 that every earthquake involves two unknowns, special to itself and shared by no other earthquake, which are very difficult to eliminate from the results. It might be thought, however, that if a sufficiently large number of observations were used the variations from one earthquake to another would be eliminated in the process of averaging, and the average would at all events yield accurate results for an ideal average earthquake, which would be a possible earthquake. The difficulty in the method, as in many other statistical investigations, arises from the fact that the observations do not constitute a fair sample.

It has been seen that the strength of the earth's crust is finite at the surface, increases to a maximum at a depth probably of the order of 100 km., and then gradually decreases with depth, probably becoming insignificant at a depth of about 400 km. Accordingly, whatever may be the cause of crustal deformation within the earth, yield will occur in the asthenosphere for much smaller stresses than are necessary to produce it in the upper parts of the crust. It has also been seen to be probable that in most crustal movements deformation in the asthenosphere probably precedes that in the upper layers. The greatest earthquakes should take place where the greatest stress is relieved by fracture, in other words at the level of greatest strength, some 100 km. below the surface. Earthquakes below that level should increase in frequency and decrease in violence with depth, while other earthquakes of moderate intensity should originate at depths between zero and 100 km.

We should expect, therefore, that the foci* of the greatest earthquakes would be at depths of the order of 100 km., though they would have a considerable range about this depth. These are the earthquakes that are recorded over the whole surface of the earth. It seldom happens, however, that a great earthquake is satisfactorily recorded at many places in its immediate vicinity, perhaps partly because it unships sensitive instruments, and partly because less sensitive instruments are often associated with bad time-keeping. The earthquakes recorded at many places near their epicentres appear to be mostly small ones, probably of small depth of focus. Thus observations of the times of earthquake waves over short

* The origin of an earthquake is often called the 'focus': the point of the surface vertically above it, the 'epicentre.'

distances refer, on the whole, to earthquakes with smaller depths of focus than those used to find the times over long distances. This systematic difference in the mean depth of focus of earthquakes, according to the distance of the observing station from the epicentre, will correspond to a systematic variation in the times of arrival of the P and S waves, which variation we have no means of eliminating.

The standard table of the times of transit of earthquake waves to various distances is that of Turner*, based on that of Zöppritz†. This gives the time-intervals between the arrivals of the P and S waves at the epicentre and at stations at distances up to 150° from the epicentre. This table is reproduced on p. 174. It must be remarked, however, that the table is so constructed as to suggest that it gives the times taken by two waves originating at one place on the earth's surface to reach other places on the surface. This suggestion, for the reason just given, is incorrect, as, of course, Prof. Turner fully recognizes. The fact that the great earthquakes most probably originate at depths of the order of 100 km. shows that the times of transit to great distances of waves originating in the surface must exceed the tabular times by quantities of the order of the times taken by a wave to travel 100 km., namely about 15 secs. for P and 25 secs. for S. Errors of a few per cent., but not more, are therefore to be expected in the times and in all inferences from them.

The following table, from the paper on the Pamir landslip already quoted, summarizes the discrepancies between the times of transit of the Pamir waves and the time-intervals over the same distances given in Turner's table. They have been grouped for ranges of 20° in epicentral distance. The time adopted for the landslip is that derived from the Tashkent observations.

Distance (degrees)	No. of observations	Mean residual of P	Mean residual of S
10– 30	7	2·3	12·2
30– 50	7	4 1	6·5
50– 70	8	36·0	12·7
70– 90	1	—	72·0
90–110	5	—	– 27·0

If the single observation between 70° and 90° is combined with the next group, we find that the mean residual of S for distances between 70° and 110° was – 10·5 secs. Thus such systematic variation with distance as exists in the residuals takes the form of an increase for P and a decrease for S. The standard deviation of the residuals for P is 25 secs., and that for S is 51 secs. Hence for means of six observations the residuals for P should be 11 secs., and for S 23 secs. The decrease for S is therefore well within the range of accidental variation, but part of the increase for P may be genuine, arising from the finite depth of focus of the earthquakes used in the tables. The absence of systematic variation for S, however,

* *Brit. Ass. Seism. Ctee. Bulletin*, May 1917.

† *Gött. Nach.* 1907, 545.

shows that the mean depth of focus of these earthquakes cannot exceed about 120 km.

Degrees	P sec.	S sec.	$S-P$ sec.	Degrees	P sec.	S sec.	$S-P$ sec.	Degrees	P sec.	S sec.	$S-P$ sec.
1	15	28	13	51	553	991	438	101	855	1565	710
2	31	55	24	52	560	1004	444	102	860	1575	715
3	47	83	36	53	566	1016	450	103	865	1584	719
4	62	110	48	54	573	1029	456	104	870	1593	723
5	77	137	60	55	579	1041	462	105	874	1602	728
6	92	164	72	56	586	1054	468	106	879	1612	733
7	106	190	84	57	592	1066	474	107	884	1621	737
8	121	217	96	58	599	1079	480	108	888	1630	742
9	136	243	107	59	605	1091	486	109	893	1639	746
10	150	269	119	60	612	1103	491	110	897	1648	751
11	164	294	130	61	619	1116	497	111	902	1657	755
12	179	319	140	62	625	1128	503	112	907	1666	759
13	193	344	151	63	632	1141	509	113	911	1674	763
14	206	368	162	64	638	1153	515	114	916	1682	766
15	219	392	173	65	645	1165	520	115	920	1690	770
16	232	415	183	66	651	1177	526	116	925	1698	773
17	245	438	193	67	658	1190	532	117	929	1706	777
18	257	460	203	68	664	1202	538	118	934	1714	780
19	269	482	213	69	671	1214	543	119	938	1722	784
20	281	503	222	70	677	1226	549	120	942	1729	787
21	293	524	231	71	683	1238	555	121	947	1737	790
22	305	545	240	72	690	1250	560	122	952	1744	792
23	317	565	248	73	696	1262	566	123	957	1752	795
24	328	584	256	74	702	1274	572	124	961	1759	798
25	338	603	265	75	709	1286	577	125	966	1766	800
26	348	622	274	76	715	1297	582	126	970	1773	803
27	358	641	283	77	721	1309	588	127	974	1780	806
28	368	659	291	78	727	1320	593	128	978	1787	809
29	378	677	299	79	733	1332	599	129	983	1794	811
30	388	694	306	80	739	1343	604	130	988	1801	813
31	398	711	313	81	745	1355	610	131	992	1807	815
32	407	728	321	82	750	1366	616	132	996	1814	818
33	416	744	328	83	756	1377	621	133	1001	1821	820
34	425	760	335	84	762	1388	626	134	1005	1827	822
35	433	775	342	85	768	1399	631	135	1009	1833	824
36	442	790	348	86	773	1410	637	136	1014	1840	826
37	450	804	354	87	779	1421	642	137	1018	1846	828
38	458	818	360	88	785	1432	647	138	1023	1852	829
39	466	832	366	89	790	1443	653	139	1027	1858	831
40	475	847	372	90	796	1454	658	140	1031	1864	833
41	483	861	378	91	801	1464	663	141	1035	1869	834
42	491	875	384	92	807	1475	668	142	1039	1875	836
43	498	888	390	93	812	1485	673	143	1043	1881	838
44	506	902	396	94	818	1496	678	144	1047	1886	839
45	513	915	402	95	823	1506	683	145	1051	1892	841
46	520	928	408	96	829	1516	687	146	1055	1897	842
47	527	941	414	97	834	1526	692	147	1059	1902	843
48	534	954	420	98	840	1536	696	148	1063	1907	844
49	540	966	426	99	845	1546	701	149	1067	1912	845
50	547	979	432	100	851	1556	705	150	1071	1917	846

12·7. *Velocities in the Interior of the Earth.* The times of transmission of seismic waves to various distances, given in this table, can be utilized to provide information about the velocities of earthquake waves over a

wide range of depth within the earth. We consider a wave spreading out from a point on the surface of the earth, and suppose its velocity in the neighbourhood of any internal point to be c, and the time taken to reach that point to be T. T will evidently be a function of the position of the point. Then the surfaces over which T is constant will mark the consecutive positions of the wave front. We shall treat the earth as spherically symmetrical, so that c is a function of the distance from the centre alone. Let r denote the distance of a point from the centre, and θ the angle between the lines joining the point and the focus to the centre. Thus r and θ are spherical polar coordinates. The wave front will at all instants be symmetrical about the line joining the centre of the earth to the focus of the earthquake. The time taken by the wave to reach a given point is

$$T = \int \frac{ds}{c}$$
$$= \int \frac{1}{c} \left\{ \left(\frac{dr}{d\theta} \right)^2 + r^2 \right\}^{\frac{1}{2}} d\theta \qquad \text{.........(1)},$$

where ds is an element of length along the path actually taken by the wave in passing from the focus to the point. The actual path is the one that makes this time the shortest possible, since the records of earthquakes give the time of the commencement of each phase. Thus the actual path has to be such as to make the integral (1) a minimum. Putting for a moment V for the integrand in (1) and ρ for $dr/d\theta$, we know from the calculus of variations that the integral is stationary if r satisfies the differential equation

$$\frac{\partial V}{\partial r} - \frac{d}{d\theta} \frac{\partial V}{\partial \rho} = 0 \qquad \text{.........(2)},$$

a first integral of which is known* to be

$$V = \rho \frac{\partial V}{\partial \rho} + p \qquad \text{.........(3)},$$

where p is a constant. On substituting for V and simplifying we find

$$\frac{r^2}{c} = p \left\{ \left(\frac{dr}{d\theta} \right)^2 + r^2 \right\}^{\frac{1}{2}} \qquad \text{.........(4)},$$

whence
$$\frac{dr}{d\theta} = \pm r \left\{ \frac{r^2}{p^2 c^2} - 1 \right\}^{\frac{1}{2}} \qquad \text{.........(5)},$$

and
$$\theta = \pm \int^r \frac{p\, dr}{r \left\{ \frac{r^2}{c^2} - p^2 \right\}^{\frac{1}{2}}} \qquad \text{.........(6)}.$$

The disturbance commences by moving inwards, so that $dr/d\theta$ is negative; thus the negative sign must be taken for the root. Again, $dr/d\theta$ vanishes when the ray reaches its nearest point to the centre. We see by (5) that

* Todhunter, *Integral Calculus*, 1883, p. 346.

the value of r at this point is pc. If χ is the corresponding value of θ, we have

$$\chi = \int_{pc}^{R} \frac{p\,dr}{r\left\{\frac{r^2}{c^2} - p^2\right\}^{\frac{1}{2}}} \quad \text{.........(7)},$$

where R is the radius of the earth. After passing this point the ray bends upwards, remaining symmetrical about the line joining the centre to the point nearest to the centre, and reaches the surface again at the point $(R, 2\chi)$. Evidently 2χ is the angular distance between the focus and the point of emergence of the ray, and may be denoted by Δ. The tables give the time taken by the ray in reaching this point as a function of Δ.

Now let P be the point (R, Δ), and T the time taken in reaching it. Let P' be a neighbouring point $(R, \Delta + d\Delta)$, and $T + dT$ the time taken in reaching it. Draw PQ perpendicular to the ray that reaches P', and let c_0 be the velocity of the wave near the surface. Then

$$P'Q = c_0 dT; \quad PP' = R d\Delta \quad \text{.........(8)},$$

and therefore

$$\frac{P'Q}{PP'} = \frac{c_0}{R}\frac{dT}{d\Delta} \quad \text{.........(9)}.$$

This ratio, however, is the cosine of the angle made with the surface by the emergent ray, and therefore, on account of the symmetrical form of the ray, is equal to the cosine of the angle made with the surface by the ray when it enters at the focus. The latter is equal to the value of $Rd\theta/ds$ when r is equal to R. But by (4)

$$\frac{d\theta}{ds} = \frac{pc}{r^2} \quad \text{.........(10)}.$$

Hence we must have

$$\frac{c_0}{R}\frac{dT}{d\Delta} = \frac{pc_0}{R},$$

or

$$p = \frac{dT}{d\Delta} \quad \text{.........(11)}.$$

Thus p is a calculable function of Δ, and therefore of χ. Therefore χ may be expressed as a function of p, $f(p)$ say. Equation (7) therefore contains only one unknown, namely c. Our problem is therefore to determine c as a function of r from this integral equation. If we put

$$\frac{r}{c} = \eta \quad \text{.........(12)},$$

equation (7) becomes

$$f(p) = p\int_{p}^{R/c_0} \frac{\frac{d}{d\eta}(\log r)\,d\eta}{(\eta^2 - p^2)^{\frac{1}{2}}} \quad \text{.........(13)},$$

which has been solved by Bateman and Herglotz*. The solution is

$$\log r = C - \frac{2}{\pi}\int_{\eta}^{R/c_0} \frac{f(p)\,dp}{(p^2 - \eta^2)^{\frac{1}{2}}} \quad \text{.........(14)},$$

* H. Bateman, *Phil. Mag.* Ser. 6, 1910, 576–87; G. Herglotz, *Phys. Zs.* 8, 1907, 145–47.

where C is a constant. When η approaches R/c_0, r approaches R, and therefore

$$C = \log R.$$

Hence finally
$$\log \frac{R}{r} = \frac{2}{\pi} \int_{\eta}^{R/c_0} \frac{f(p)\, dp}{(p^2 - \eta^2)^{\frac{1}{2}}} \qquad \text{.........(15).}$$

This equation gives r as a function of η, and on eliminating η between this and (12) we obtain c as a function of r. The calculation, which is extremely laborious, has been carried out by Prof. C. G. Knott*. His results are as follows:

Primary wave		Secondary wave	
r (km.)	c_1 (km./sec.)	r (km.)	c_2 (km./sec.)
6378	7·18	6378	3·98
6349	7·38	6348	4·10
6307	7·59	6298	4·22
6252	7·80	6251	4 37
6194	8·00	6183	4·50
6142	8·24	6122	4·65
6075	8·48	6045	4·79
6010	8·74	5970	4·97
5940	9·01	5887	5·14
5867	9·31	5801†	5·32
5783	9·61	5705	5·53
5705	9·96	5618†	5·77
5612	10·32	5488	5·98
5501	10·66	5361	6·24
5389	11·07	5216	6·50
5263	11·49	5042	6 77
5115	11·90	4929	6·88
4968	12·39	4716	6·85
4794	12·89	4306	6·84
4724	12·87	3920	6·84
4534	12·77	3536	6·85
4405	12·85	3139	6·85
4044	12·84	2703	6·74
3676	12·81	2352	6·84
3268	12·67	1929	6·72
2888	12·73		

These values are shown graphically in Fig. 6. It is seen that both velocities increase almost as linear functions of the depth, down to a depth of about 1600 km., and below that are practically constant. To assume that these two relations hold exactly would only introduce errors of a few per cent., comparable with those already introduced by the assumption that the foci of the earthquakes are in the surface. The ratio of two velocities, again, hardly varies from 1·8. As is well known, Poisson's theory of the solid state implies that the two properties of a solid denoted by λ and μ are equal, and therefore, since the velocities of the two waves are $\{(\lambda + 2\mu)/\rho\}^{\frac{1}{2}}$ and $\{\mu/\rho\}^{\frac{1}{2}}$, their ratio should be $3^{\frac{1}{2}}$, or 1·73.

Earthquake waves are not often recorded at distances greater than 140°, and it is for this reason that Knott has been unable to extend his calculations to greater depths than those given in the above table. It

* *Proc. Roy. Soc. Edin.* **39**, 1919, 158–208.

† Corrections to Knott's values found by Mr R. Stoneley and acknowledged by Knott.

appears probable that the material of the earth at still greater depths is imperfectly elastic in such a way as to prevent the passage of seismic waves. The alternative possibility, however, that some increase in the velocity of propagation at these depths causes seismic waves to be totally reflected instead of penetrating them, appears to merit some consideration.

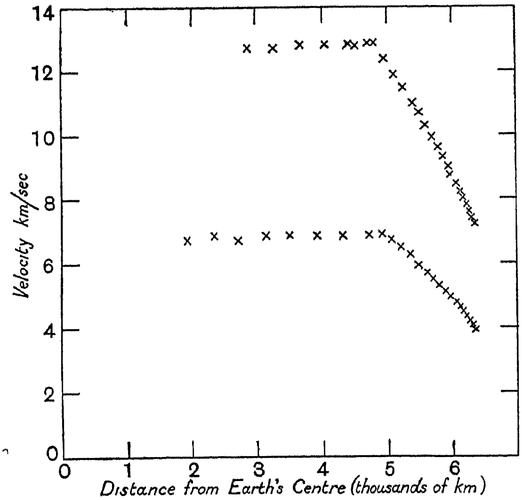


Fig 6.

12·8. *Variations of Composition of Rocks near the Surface.* It will be noticed that the velocities of propagation of earthquake waves near the surface, as found by Knott, differ considerably from those found by Dr Wrinch and myself for the waves from the Oppau explosion. The discrepancy may be attributed in part to the depths of the foci of the earthquakes used in the tables. However, the velocity of the primary wave has increased only to 7·8 km./sec., according to Knott's results, at a depth of 126 km., and a wave penetrating to that depth, according to Table IV of Knott's paper, emerges at a distance of 1200 km. from the epicentre; but the time of transit to this distance is 164 secs., while the error due to depth of focus can be only a few seconds. Hence we may suppose that Knott's result for this depth is unlikely to be in error by more than a few per cent. Thus there must be a considerable change in the velocity of propagation of earthquake waves within the first 126 km. from the surface. Such a change has already been inferred from the apparent existence of Love waves; and now that we have approximate estimates of the velocities in the respective layers, we can use the long wave phase to give some idea of the depth of the layer where the waves travel with the velocities inferred from the Oppau explosion. The velocity of the long waves was found by O. Klotz*, from the mean of a large number of

* *Bull. Seis. Soc. Amer.* 7, 1907, 67–71.

observations, to be about 3·8 km./sec. The reliability of this estimate must not be over-emphasized, in view of the great range of real variation in the velocity of propagation of the long waves. In the absence of the elaborate analysis of the long wave phase that would be necessary for a completely satisfactory discussion, we may adopt this estimate. The velocity of distortional waves found from the Oppau explosion was 3·1 km./sec., decidedly less than the velocity of typical long waves. Neither Rayleigh nor Love waves of wave length short compared to the depth of the surface layer could have a velocity greater than that of distortional waves in the surface layer, and therefore the long waves must extend well into the layer of greater velocity. The velocity of long Rayleigh waves, being about 0·92 of that of distortional waves in the under layer, namely 3·98 km./sec., and that of Love waves, being equal to that of these latter waves, agree well with the observed velocity of 3·8 km./sec. for the long waves. Now the prevailing period of the long waves is about 15 secs. This, combined with their velocity, implies a wave length of 57 km. In consequence of the relation $s = 0{\cdot}39\kappa$, given on p. 159, this shows that the amplitude of these waves must be inappreciable at a depth of 150 km.; and in order that the velocity may be determined mainly by the lower layer, it is therefore necessary that the depth of the upper layer should be a small fraction of this, probably not exceeding 30 km.

The Nördlingen record of the Oppau explosion affords a further datum. It showed a well marked long wave phase, commencing just after S. The velocity of these long waves was therefore almost equal to that of distortional waves in the upper layer, and they may be supposed to have been almost confined to this layer. Their period was 2 to 3 secs., implying a wave length of 6 to 9 km. Hence the vertical extent of the layer to which they were practically confined must have been at least of the order of 10 km. Combining this result with that of the last paragraph, we see that the thickness of the layer that carried the Oppau waves was probably more than 10 km., and probably less than 30 km. This corresponds well with the inference of 6·51 that the thickness of the granitic layer of the continents is of the order of 15 km., and suggests that the Oppau waves have afforded us an estimate of the velocities of earthquake waves in the granitic layer. If so, the estimates obtained by Knott for surface rocks must correspond to the basic rocks below the granitic layer.

The long waves continued at Nördlingen until their period had fallen to 1·7 secs. The velocity corresponding to the time of arrival of these waves was only 0·76 km./sec., so that a medium only 3 km. thick, and with a distortional wave velocity of 0·8 km./sec., would have sufficed to carry them. Similar waves were recorded at Strasbourg. It is natural to suppose that these waves travelled in the sedimentary rocks that overlie the granite.

12·81. The elastic constants of a number of rocks have been determined by Adams and Coker*, and it is desirable to compare them with those just obtained, in order to check the suggestions made regarding the nature of the various layers. The following velocities are obtained from their work, the elastic constants having been translated into the notation used in this book, and from data kindly provided by Mr A. Harker concerning the densities of the rocks. C.G.S. units are used.

Rock	$\lambda+2\mu$	μ	ρ	c_1	c_2
Baveno granite	$5{\cdot}678 \times 10^{11}$	$1{\cdot}875 \times 10^{11}$	2·72	$4{\cdot}67 \times 10^5$	$2{\cdot}62 \times 10^5$
Peterhead granite	$6{\cdot}420 \times 10^{11}$	$2{\cdot}340 \times 10^{11}$	2·69	$4{\cdot}88 \times 10^5$	$2{\cdot}95 \times 10^5$
Westerly granite	$5{\cdot}792 \times 10^{11}$	$2{\cdot}080 \times 10^{11}$	2·64	$4{\cdot}68 \times 10^5$	$2{\cdot}81 \times 10^5$
Quincey granite (1)	$5{\cdot}305 \times 10^{11}$	$1{\cdot}916 \times 10^{11}$	2·66	$4{\cdot}46 \times 10^5$	$2{\cdot}69 \times 10^5$
Quincey granite (2)	$6{\cdot}304 \times 10^{11}$	$2{\cdot}373 \times 10^{11}$	2·66	$4{\cdot}87 \times 10^5$	$2{\cdot}99 \times 10^5$
Stanstead granite	$4{\cdot}793 \times 10^{11}$	$1{\cdot}556 \times 10^{11}$	2·66	$4{\cdot}25 \times 10^5$	$2{\cdot}42 \times 10^5$
New Glasgow anorthosite	$10{\cdot}127 \times 10^{11}$	$3{\cdot}275 \times 10^{11}$	2·65–2·67	$6{\cdot}18 \times 10^5$	$3{\cdot}51 \times 10^5$
New Glasgow gabbro	$12{\cdot}429 \times 10^{11}$	$4{\cdot}380 \times 10^{11}$	—	—	—
Sudbury diabase	$12{\cdot}262 \times 10^{11}$	$3{\cdot}700 \times 10^{11}$	3·07	$6{\cdot}39 \times 10^5$	$3{\cdot}51 \times 10^5$
Ohio sandstone	$2{\cdot}066 \times 10^{11}$	$0{\cdot}612 \times 10^{11}$	2–2·2	3×10^5	$1{\cdot}7 \times 10^5$

The Oppau velocities are seen to be slightly greater than those for the granites, the excess being attributable to the influence of pressure in increasing the elastic constants of materials, but they are much less than those for the basic rocks. Similarly Knott's surface velocities somewhat exceed those for the basic rocks, the difference being similarly explicable. Accordingly the identification of the layer transmitting the Oppau waves with the granitic layer, and of the uppermost layer considered by Knott with the basic layer immediately below, are consistent with laboratory measurements. The velocity attributed to distortional waves in the sedimentary layer is less than that in Ohio sandstone, but may correspond to some sedimentary rock that has been less firmly consolidated.

12·82. The transition from the basic to the granitic layer must have an important influence on the angles of emergence of earthquake waves. A wave from the interior at grazing incidence would be refracted on entering the granitic layer, and would emerge at an angle of about 40° to the horizon. Thus there is a definite minimum possible angle of emergence at the surface if the base of the granitic layer is horizontal.

In the case of a superficial shock, the P and S waves through the sedimentary layer would arrive after the onset of the long waves through the granitic layer, and would therefore be swamped by them. The S wave through the granitic layer, again, travels more slowly than Rayleigh waves of period 10–20 secs., and could be detected in the Oppau waves only because the superficial character of the shock and the comparative nearness of the observing stations prevented these long Rayleigh waves from being developed before arrival. In addition to the direct P and S waves through the sedimentary layer, and to those refracted into the granitic layer and spreading out in this, it is possible for these waves to penetrate into the

* *Amer. Journ. of Science*, Ser. 4, 22, 1906, 95–123.

basic layer, spread out in it, and be refracted up again to the observing station, and it is possible that the greater speed of the waves in the basic layer may be enough to make them travel this path in a shorter time than the corresponding path through the granitic layer. Yet the records of the Oppau explosion showed no trace of these waves. The reason appears to be that in refraction the wave that reaches the basic layer is so much spread out that its amplitude becomes inappreciable. In a rough estimate made in the paper on the Oppau explosion already mentioned, we inferred that the amplitude of the wave through the basic layer, as recorded at München, could theoretically not have exceeded a tenth of that of the direct wave, and the latter was itself only just observable.

The wave reflected at the interface would be insignificant for the same reason.

It may be asked whether the spreading out of the energy would not equally affect the passage of the wave from the sedimentary layer to the granitic one, in which case the wave in the latter might be as much affected as that in the basic layer. But the truth of the laws of refraction depends on the condition that each vibrating layer shall be several wave lengths in thickness. The sedimentary layer does not satisfy this condition. Hence an impulse applied to it will produce almost the same disturbance inside as if it was applied directly to the granitic layer.

12·9. It has been seen that, if the data afforded by earthquakes were supplemented by knowledge of the times of the shocks and the depths of the foci, a great improvement in our knowledge of seismology would result. The needful information appears to be, first, more knowledge of the behaviour of waves in the surface layer in other parts of the earth, and especially under the ocean, and second, a good means of finding the depth of focus.

12·91. The velocities of waves in the surface layer could be surveyed very fully by means of artificial explosions. At Oppau 4500 tons of explosive were destroyed*, but only a small fraction of the energy of the explosion can have gone into the earth. The impulses upwards on the air and downwards on the ground must, by Newton's third law, have been equal. Hence the energies imparted to the air and the earth must have been in the ratio of the initial vertical velocities, which must have been thousands of times greater for the air than for the ground. Thus much less than a thousandth of the energy of the explosion can have entered the earth; the rest went into the sound wave. Now if an artificial explosion were made in an underground chamber, tightly packed with explosive, practically the whole of the energy would go into the seismic wave. Records as intense as those of Oppau and at similar distances could therefore be obtained by the use of a few tons of explosive.

* *Nature*, **108**, 1921, 278.

12·92. Three methods appear likely to be of service in finding the depths of foci. In the case of a sudden earthquake, the interval between P and S gives valuable information about the distance of the shock, and if such information for neighbouring stations were available for many earthquakes, the uncertainty arising from the possible variation of velocity in the upper layers could be eliminated.

The second method, due to von Seebach*, is as follows. If c be the velocity of the wave, d the depth of the focus, supposed small in comparison with the radius of the earth, and x the distance of the station from the epicentre, the time T taken in transit is given by

$$cT = (x^2 + d^2)^{\frac{1}{2}} \quad \text{or} \quad c^2T^2 - x^2 = d^2.$$

Thus if the values of x and cT for several observing stations are plotted on a graph, the resulting points will lie on a hyperbola, whose major semi-axis is the depth of focus.

Prof. H. H. Turner has introduced a further method†, which may be useful in comparing different earthquakes, but does not give the depth of focus of any single earthquake absolutely. His method is based on a study of the difference between the times of arrival of the waves at stations respectively near the epicentre and near its antipodes. This difference should be closely correlated with the depth of focus, and if the depth of focus of one earthquake were known so well that it could be used as a standard, Turner's method would determine all others. So far, however, it has suggested that the mean depth of the foci of the earthquakes used in the tables is about 200 km., greater than the comparison with the Pamir landslip would appear to allow; the disagreement remains at present unexplained‡.

* C. Davison, *Manual of Seismology*, 1921, 127.

† *M.N.R.A.S.*, *Geophys. Suppt.* 1, 1922, 1–13; 1923, 50–55.

‡ But Mr R. D. Oldham brings further evidence tending to show the Pamir earthquake to have been deep-seated. See *Q. J. Geol. Soc.*, 79, 1923, 231–45.

CHAPTER XIII

The Figures of the Earth and Moon

"There's mair ways o' killin' a pig than by greasin' him wi' het butter." Northumbrian proverb.

13·1. We have so far considered the earth as a stationary body under no external forces, evolving under the influence of the changes in its internal temperature, controlled by the mutual gravitation of its parts. In this chapter inequalities produced by slow rotation and small, slowly varying tidal disturbances will be considered. Since the departures of the earth and the moon from spherical symmetry are small, it is legitimate to discuss the two types of disturbance separately, and to regard the effect of the two together as the sum of the effects of the two separately; thus the effects of rotation and tides are simply superposed on the mountains, continents, and other inequalities already discussed.

13·11. *The Gravitational Potential of a nearly Spherical Body.* We shall evidently need to know the gravitational forces due to the disturbed body itself. Let us consider first the potential due to a homogeneous mass of almost spherical form. Its density is to be ρ, supposed uniform. Take spherical polar coordinates (r, θ, ϕ) with regard to a point near the centre. The radius vector to any point on the surface can be expressed in the form

$$r = a\,(1 + Y_1 + Y_2 + \dots + Y_n + \dots) \qquad \dots\dots\dots(1),$$

where the Y's are spherical surface harmonic functions of θ and ϕ, the angular coordinates of the point on the surface. It will be supposed that the mass is so nearly spherical that all the Y's are small enough for their squares and products to be neglected. If the gravitational potential outside be U_0, and that inside be U_1, the conditions for the continuity of the potential and its normal derivative are easily seen to be satisfied, subject to the above proviso concerning the neglect of squares and products of the Y's, if

$$U_0 = \frac{4}{3}\pi f\rho a^3\left(\frac{1}{r} + \frac{aY_1'}{r^2} + \frac{3}{5}\frac{a^2Y_2'}{r^3} + \dots + \frac{3}{2n+1}\frac{a^nY_n'}{r^{n+1}} + \dots\right) \qquad \dots\dots\dots(2),$$

$$U_1 = \frac{4}{3}\pi f\rho a^3\left(\frac{3a^2 - r^2}{2a^3} + \frac{rY_1'}{a^2} + \frac{3}{5}\frac{r^2Y_2'}{a^3} + \dots + \frac{3}{2n+1}\frac{r^nY_n'}{a^{n+1}} + \dots\right) \dots(3),$$

where the Y''s are the same functions of θ' and ϕ' that the Y's are of θ and ϕ, and f is the constant of gravitation. U_1 and U_0 are therefore the appropriate values of the internal and external potentials.

13·12. Passing now to the case of a heterogeneous body, let the equation of the stratum whose density is ρ be

$$r = r_1\,(1 + Y_1 + Y_2 + \dots + Y_n) \qquad \dots\dots\dots(1),$$

where the Y's may evidently be functions of r_1. Consider the potential due to the shell between the layers where the densities are respectively ρ' and $\rho' + d\rho'$. To the first degree in $d\rho'$, the contribution to U_0 due to this shell is the difference between the potentials due to two homogeneous bodies, both of density ρ', one filling the outer of these layers and the other the inner. Then the potential due to a homogeneous body filling either stratum can be found from 13·11 (2) and (3), and the potential due to the layer between the two strata considered is

$$\frac{4}{3}\pi f\rho' \frac{\partial}{\partial a'}\left\{\frac{a'^3}{r} + \frac{a'^4 Y_1'}{r^2} + \frac{3}{5}\frac{a'^5 Y_2'}{r^3} + \dots + \frac{3}{2n+1}\frac{a'^{n+3}Y_n'}{r^{n+1}} + \dots\right\} da' \quad\text{.........(2)},$$

where a' is the value of r_1 corresponding to $\rho = \rho'$, and other accented letters also refer to this layer. By integrating for all layers we may therefore find that the potential at external points due to the whole heterogeneous body is

$$U_0 = \frac{4}{3}\pi f\int_0^a \rho' \frac{\partial}{\partial a'}\left\{\frac{a'^3}{r} + \frac{a'^4 Y_1'}{r^2} + \frac{3}{5}\frac{a'^5 Y_2'}{r^3} + \dots + \frac{3}{2n+1}\frac{a'^{n+3}Y_n'}{r^{n+1}} + \dots\right\} da' \quad\text{.........(3)},$$

where a is the mean radius of the outer surface of the body. We may find the potential at an internal point similarly. If the mean radius of the stratum of equal density through the point is r_1, the matter within this contributes to the potential an amount

$$\frac{4}{3}\pi f\int_0^{r_1} \rho' \frac{\partial}{\partial a'}\left\{\frac{a'^3}{r} + \frac{a'^4 Y_1'}{r^2} + \frac{3}{5}\frac{a'^5 Y_2'}{r^3} + \dots + \frac{3}{2n+1}\frac{a'^{n+3}Y_n'}{r^{n+1}} + \dots\right\} da' \quad\text{...(4)},$$

while the matter outside of it is found from 13·11 (3) in a similar way to give

$$\frac{4}{3}\pi f\int_{r_1}^{a} \rho' \frac{\partial}{\partial a'}\left\{\frac{3}{2}a'^2 + a' r Y_1' + \frac{3}{5}r^2 Y_2' + \dots + \frac{3}{2n+1}\frac{r^n}{a'^{n-2}}Y_n' + \dots\right\} da' \quad\text{.........(5)}.$$

Thus U_1 is the sum of the two expressions (4) and (5).

13·13. The mass of the heterogeneous body is evidently given by

$$M = 4\pi\int_0^a \rho' a'^2 da' \quad\text{.........(1)}.$$

We shall frequently have occasion to use the function

$$S(r_1) = 3\int_0^{r_1} \rho' a'^2 da' \quad\text{.........(2)}.$$

Thus $S(r_1)$ is $3/4\pi$ times the mass within the stratum of equal density whose mean radius is r_1. In particular,

$$S(a) = \frac{3M}{4\pi} \quad\text{.........(3)}.$$

We shall also need to use the moments of inertia of the body and the differences between them. If A, B, C denote the moments of inertia about the axes of x, y, and z respectively, we have

$$A = \iiint \rho\,(y^2 + z^2)\,d\tau \quad\text{.........(4)},$$

with two similar expressions, each triple integral being taken through the body By subtraction,

$$C - A = \iiint \rho\, (x^2 - z^2)\, d\tau \qquad \text{.........(5)},$$

with two similar expressions. Now let us put

$$x = r \sin\theta \cos\chi \qquad \text{.........(6)},$$

$$y = r \sin\theta \sin\chi \qquad \text{.........(7)},$$

$$z = r \cos\theta \qquad \text{.........(8)}.$$

Then

$$C = \int_0^R \int_0^\pi \int_0^{2\pi} \rho r^4 \sin^3\theta\, dr\, d\theta\, d\chi \qquad \text{.........(9)},$$

where R denotes the distance from the centre to the surface in the direction given by (θ, χ). Suppose first that the body is homogeneous. We can write

$$\sin^2\theta = \tfrac{2}{3} + (\tfrac{1}{3} - \cos^2\theta) \qquad \text{.........(10)},$$

and the second part of this is a spherical surface harmonic of order 2. Then

$$C = \tfrac{1}{5} \int_0^\pi \int_0^{2\pi} \rho R^5 \{\tfrac{2}{3} + (\tfrac{1}{3} - \cos^2\theta)\} \sin\theta\, d\theta\, d\chi \quad \text{.........(11)},$$

and on substitution for R from 13·11 (1) and neglecting squares of the Y's,

$$C = \tfrac{1}{5} \int_0^\pi \int_0^{2\pi} \rho a^5 (1 + 5Y_1 + 5Y_2 + \ldots + 5Y_n) \{\tfrac{2}{3} + (\tfrac{1}{3} - \cos^2\theta)\} \sin\theta\, d\theta\, d\chi$$

$$\text{.........(12)}.$$

Now if S_m and S_n be any two surface harmonics of different orders we know that

$$\int_0^\pi \int_0^{2\pi} S_m S_n \sin\theta\, d\theta\, d\phi = 0,$$

and therefore when the integrand in (12) is multiplied out and integrated, the only terms that do not vanish are

$$C = \tfrac{8}{15}\pi\rho a^5 + \rho a^5 \int_0^\pi \int_0^{2\pi} Y_2 (\tfrac{1}{3} - \cos^2\theta) \sin\theta\, d\theta\, d\phi \quad \text{.........(13)}.$$

The harmonics of orders different from 0 and 2 therefore contribute nothing to the moments of inertia.

Now Y_2 must be of the form

$$Y_2 = (\alpha x^2 + \beta y^2 + \gamma z^2 + 2fyz + 2gzx + 2hxy)/r^2 \qquad \text{...(14)}.$$

When this is substituted in (13) and integrated, the product terms contain factors

$$\int_0^{2\pi} \sin\phi\, d\phi, \quad \int_0^{2\pi} \cos\phi\, d\phi, \quad \int_0^{2\pi} \sin\phi \cos\phi\, d\phi,$$

all of which vanish. The other terms in C therefore become

$$C = \tfrac{8}{15}\pi\rho a^5 + \tfrac{8}{45}\pi\rho a^5 (\alpha + \beta - 2\gamma) \qquad \text{.........(15)}.$$

If then we retain only the largest terms, we have

$$C = \tfrac{8}{15}\pi\rho a^5 \qquad \text{.........(16)},$$

$$C - A = \tfrac{8}{15}\pi\rho a^5 (\alpha - \gamma) \qquad \text{.........(17)},$$

with symmetrical expressions for the other moments of inertia and their differences. It follows that for a homogeneous ellipsoid, if the axes are the

principal axes, the ratio $(C - A)/C$ is equal to the ellipticity of the section by the plane $y = 0$.

The two results (16) and (17) can be generalized by the method of 13·12. We thus get

$$C = \tfrac{8}{15}\pi \int_0^a \rho' da'^5 \quad \text{.........(18)},$$

$$C - A = \tfrac{8}{15}\pi \int_0^a \rho' d\{a'^5 (\alpha - \gamma)\} \quad \text{.........(19)}.$$

Thus the ratio $(C - A)/C$ for a heterogeneous body is a weighted mean of the ellipticities of the strata of equal density.

13·2. *Condition for Equilibrium of a Rotating Fluid.* Consider a body in a state of steady rotation about a fixed axis, which will be taken to be the axis of z. Let the angular velocity be Ω. Then the component accelerations of any element of the body are $(-\Omega^2 x, -\Omega^2 y, 0)$. If now any part of the body is under no shearing stress (in other words, if it is fluid, or if it is a solid, but has had any stress-differences removed by plastic adjustment) the equations of motion will be, as in ordinary hydrodynamics,

$$-\Omega^2 x = \frac{\partial U}{\partial x} - \frac{1}{\rho}\frac{\partial p}{\partial x} \quad \text{.........(1)},$$

$$-\Omega^2 y = \frac{\partial U}{\partial y} - \frac{1}{\rho}\frac{\partial p}{\partial y} \quad \text{.........(2)},$$

$$0 = \frac{\partial U}{\partial z} - \frac{1}{\rho}\frac{\partial p}{\partial z} \quad \text{.........(3)},$$

where U is the complete gravitation potential, p the pressure, and ρ the density. If we write

$$\Psi = U + \tfrac{1}{2}\Omega^2 (x^2 + y^2) \quad \text{.........(4)},$$

the equations of motion are together equivalent to the single total differential equation

$$dp = \rho d\Psi \quad \text{.........(5)}.$$

It follows that p is a function of Ψ, and that ρ is either the same everywhere or another function of Ψ. In particular, if the fluid or quasi-fluid has a free surface, the pressure is uniformly zero over this surface, and therefore Ψ is constant over it.

It has been seen that the whole of the interior of the earth, at depths greater than 400 km., is probably unable to withstand even small shearing stresses for long periods. The whole of the matter below this level must therefore be still in a hydrostatic state, while just after solidification this state must have held right up to the surface. Accordingly the theory just given was applicable to the whole of the earth just after solidification, and is still applicable to most of the interior. It is certain, however, that changes occurring in the crust at depths less than 400 km. have prevented the hydrostatic state from persisting at these depths, and it is possible that such changes may have influenced the adjustment of the form of

the outer layers to the earth's actual speed of rotation. The discrepancy is unlikely to be large, since the thickness of these layers is only a small fraction of the radius of the earth, and they would probably bend in such a way as to adjust themselves with some accuracy to inequalities of great horizontal extent, such as the one produced by rotation; but it may be appreciable.

On the other hand, we know that the ocean is a fluid and, apart from the disturbing influences of gravitational and meteorological tides, which are small in comparison with its ellipticity of figure, the theory must hold exactly. It is therefore desirable to make what progress we can from empirical data obtainable on the surface of the ocean, before we proceed to supplement them by means of a theory of the conditions in the interior of the earth.

13·21. We notice that in general we can write

$$\tfrac{1}{2}\Omega^2 (x^2 + y^2) = \tfrac{1}{3}\Omega^2 r^2 + \tfrac{1}{2}\Omega^2 r^2 (\tfrac{1}{3} - \cos^2\theta) \qquad \text{.........(1)},$$

where θ is the colatitude, or the difference between the true latitude and 90°. Thus the effect of rotation can be expressed by the addition to the gravitation potential of two small terms, one symmetrical about the centre of the body, and the other a solid harmonic of order 2.

13·3. *The Form of the Ocean Surface.* Let us now proceed to consider the effect of rotation on the form of the ocean surface. We have seen that it may be represented by introducing a disturbing potential involving terms of orders 0 and 2. If now we consider only terms of these orders, and use the suffix a to denote the surface value of the quantity concerned, we have from 13·12 (3)

$$U_0 = \frac{4}{3}\pi f \int_0^a \rho' \frac{\partial}{\partial a'}\left(\frac{a'^3}{r} + \frac{3}{5}\frac{a'^5}{r^3} Y_2'\right) da' \qquad \text{.........(1)},$$

and the form of the surface is given by

$$r = a(1 + Y_{2,a}) \qquad \text{.........(2)}.$$

Substituting from (2) in (1), and neglecting terms of the second degree in Y_2, we have

$$U_{0,a} = 4\frac{\pi f}{a}\int_0^a \rho' a'^2 da' (1 - Y_{2,a}) + \frac{4}{5}\frac{\pi f}{a^3}\int_0^a \rho' \frac{\partial}{\partial a'}(a'^5 Y_2')\, da' \qquad \text{......(3)},$$

and on substituting in Ψ, and remembering 13·13 (1), we have

$$\frac{fM}{a}(1 - Y_{2,a}) + \frac{4}{5}\frac{\pi f}{a^3}\int_0^a \rho' \frac{\partial}{\partial a'}(a'^5 Y_2')\, da' + \tfrac{1}{3}\Omega^2 a^2 + \tfrac{1}{2}\Omega^2 a^2 (\tfrac{1}{3} - \cos^2\theta)$$
$$= \text{constant} \qquad \text{.........(4)}.$$

The terms fM/a and $\frac{1}{3}\Omega^2 a^2$ are constant over the surface as they stand. Now Y_2' will in general be a linear function of the five different surface harmonics of the second order. Thus (4) gives, on equating coefficients of each harmonic separately, one equation satisfied by each coefficient. At present, however, we are concerned only with the harmonic $\frac{1}{3} - \cos^2\theta$.

If its coefficient in Y_2 is ϵ, we have on picking out coefficients of $\frac{1}{3} - \cos^2\theta$ in (4)

$$-\frac{fM}{a}\epsilon_a + \frac{4}{5}\frac{\pi f}{a^3}\int_0^a \rho' \frac{\partial}{\partial a'}(a'^5\epsilon')\,da' + \tfrac{1}{2}\Omega^2 a^2 = 0 \quad \text{.........(5)}.$$

Now if we ignore all other inequalities, so that the surface is

$$r = a\{1 + \epsilon_a(\tfrac{1}{3} - \cos^2\theta)\} \quad \text{.........(6)},$$

we see that the ratio of the equatorial and polar radii of the surface is, to the first degree in ϵ_a, equal to $1 - \epsilon_a$. Thus ϵ_a is the ellipticity of the ocean surface. Similarly ϵ is the ellipticity of any other stratum of equal density. Now using 13·13 (19), we have

$$-\frac{fM}{a}\epsilon_a + \frac{3}{2}\frac{f(C-A)}{a^3} + \tfrac{1}{2}\Omega^2 a^2 = 0 \quad \text{.........(7)}.$$

Let us put

$$\Omega^2 a^3/fM = m \quad \text{.........(8)},$$

so that m is a small number. Then (7) becomes

$$\frac{3}{2}\frac{C-A}{Ma^2} = \epsilon_a - \tfrac{1}{2}m \quad \text{.........(9)}.$$

This important result shows that if the ellipticity of the ocean surface and the ratio of the centrifugal force at the equator to mean gravity are known, it is possible to infer the difference between the principal moments of inertia of the earth.

Referring back to (1), we see that the relevant part of U_0 is

$$U_0 = \frac{fM}{r} + \frac{3}{2}\frac{f(C-A)}{r^3}(\tfrac{1}{3} - \cos^2\theta) \quad \text{.........(10)}.$$

Now the observable value of gravity is the acceleration of a freely falling body with regard to the crust below it. The acceleration relative to the centre of the earth is $\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) U_0$; but each particle of the earth has, in consequence of the rotation, an acceleration $(-\Omega^2 x, -\Omega^2 y, 0)$. The acceleration of a falling body with regard to the earth below it is therefore

$$\left(\frac{\partial U_0}{\partial x} + \Omega^2 x, \frac{\partial U_0}{\partial y} + \Omega^2 y, \frac{\partial U_0}{\partial z}\right) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\Psi \quad \text{......(11)},$$

by 13·2 (4). Now Ψ is constant over the ocean surface, and therefore its gradient has no component along the surface. Thus observable gravity is normal to the ocean surface. If its value is g,

$$\begin{aligned} g^2 &= \left(\frac{\partial\Psi}{\partial x}\right)^2 + \left(\frac{\partial\Psi}{\partial y}\right)^2 + \left(\frac{\partial\Psi}{\partial z}\right)^2 \\ &= \left(\frac{\partial\Psi}{\partial r}\right)^2 + \left(\frac{\partial\Psi}{r\,\partial\theta}\right)^2 \quad \text{.........(12)}. \end{aligned}$$

Now $\partial\Psi/r\partial\theta$ is a small quantity, and therefore we may neglect its square and put

$$g = -\frac{\partial\Psi}{\partial r} \quad \text{.........(13)},$$

the negative sign indicating that it is measured inwards. Now from (9) and (10), and 13·21 (1),

$$\Upsilon = fM\left[\frac{1}{r} + \frac{a^2}{r^3}(\epsilon_a - \tfrac{1}{2}m)(\tfrac{1}{3} - \cos^2\theta) + \frac{mr^2}{a^3}\{\tfrac{1}{3} + \tfrac{1}{2}(\tfrac{1}{3} - \cos^2\theta)\}\right] \quad \ldots(14),$$

and therefore

$$g = fM\left[\frac{1}{r^2} + \frac{3a^2}{r^4}(\epsilon_a - \tfrac{1}{2}m)(\tfrac{1}{3} - \cos^2\theta) - \frac{2mr}{a^3}\{\tfrac{1}{3} + \tfrac{1}{2}(\tfrac{1}{3} - \cos^2\theta)\}\right] \ldots(15),$$

and on substituting for r from (6) and neglecting squares of ϵ_a and m,

$$g = \frac{fM}{a^2}\{(1 - \tfrac{2}{3}m) - (\tfrac{5}{2}m - \epsilon_a)(\tfrac{1}{3} - \cos^2\theta)\}$$

$$= \frac{fM}{a^2}\{1 + \alpha + (\tfrac{5}{2}m - \epsilon_a)\cos^2\theta\} \quad \ldots\ldots\ldots(16),$$

where α is a small quantity independent of θ. If now G denote the intensity of gravity on the equator, where $\theta = \frac{1}{2}\pi$, we evidently must have

$$G = \frac{fM}{a^2}(1 + \alpha) \quad \ldots\ldots\ldots(17),$$

and on dividing (16) by (17), and again rejecting squares of small quantities,

$$g = G\{1 + (\tfrac{5}{2}m - \epsilon_a)\cos^2\theta\} \quad \ldots\ldots\ldots(18).$$

Thus the increase of gravity in passing from the equator to some other latitude is proportional to the square of the sine of the latitude reached. The formula (18) is due to Clairaut. It will be seen that if we know m, and can find by experiment the variation of gravity over a wide range of latitude, it will enable us to find ϵ independently of trigonometric determinations.

From a comparison of gravity determinations in many widely separated regions Bowie* has found that

$$\tfrac{5}{2}m - \epsilon_a = 0{\cdot}005294 \quad \ldots\ldots\ldots(19).$$

Also $$1/m = 288{\cdot}4 \quad \ldots\ldots\ldots(20).$$

Thus $$\epsilon_a = 0{\cdot}003373 = 1/296{\cdot}4 \quad \ldots\ldots\ldots(21).$$

Bowie gets $$\epsilon_a = 1/297{\cdot}4 \quad \ldots\ldots\ldots(22),$$

apparently by allowing for the second degree terms which have here been neglected. For the sake of internal consistency, however, the value (21) will be used in the following calculation. Now (9) gives

$$\frac{3}{2}\frac{C - A}{Ma^2} = 0{\cdot}001640 \quad \ldots\ldots\ldots(23).$$

The ratio $(C - A)/C$ can be found from the period of the precession of the equinoxes, and is known with a smaller probable error than the surface ellipticity. We have

$$\frac{C - A}{C} = \frac{1}{305{\cdot}6} = 0{\cdot}003272 \quad \ldots\ldots\ldots(24).$$

* *U.S. Coast and Geodetic Survey, Spec. Publ.* **40, 1917, 134.**

On dividing (23) by (24) we have

$$\frac{C}{Ma^2} = 0{\cdot}3341 \qquad \text{.........(25).}$$

If the earth were exactly homogeneous, this ratio would be 0·4000. Thus we have a definite indication that the earth is denser near the centre than outside.

13·4. *The Condition for Absence of Shearing Stress Internally.* The above results depend only on the hypothesis that there is no shearing stress in the ocean. If now we introduce the further hypothesis, which we have seen in 13·2 is unlikely to be seriously in error, that, so far as inequalities of large extent are concerned, the earth is in a hydrostatic state throughout, we see that Ψ must be constant over all the surfaces of equal density within the earth. Hence if in it we substitute for r its value

$$r_1 (1 + Y_1 + Y_2 + \ldots + Y_n + \ldots) \qquad \text{.........(1)}$$

from 13·12 (1), Ψ must reduce to a function of r_1 alone. If this is done, and we neglect squares and products of the Y's, the terms of order n give

$$-\frac{Y_n}{r_1}\int_0^{r_1}\rho' a'^2 da' + \frac{1}{(2n+1)\, r_1^{n+1}}\int_0^{r_1}\rho' \frac{\partial}{\partial a'}(a'^{n+3} Y_n')\, da' + \frac{r_1^n}{2n+1}\int_{r_1}^{a}\rho' \frac{\partial}{\partial a'}\left(\frac{Y_n'}{a'^{n-2}}\right) da' = 0 \qquad \text{.........(2),}$$

except for $n = 0$, when the condition is satisfied automatically, and for $n = 2$, when the right side has to be replaced by

$$-\frac{1}{8\pi f}\Omega^2 r_1^2 \left(\tfrac{1}{3} - \cos^2\theta\right).$$

We have thus a separate linear integral equation for each Y except Y_0.

Let us first consider the harmonic corresponding to the polar flattening, so that Y_2 is equal to $\epsilon\left(\frac{1}{3} - \cos^2\theta\right)$, where the ellipticity ϵ is a small quantity and a function of r_1 alone. Then the surfaces of equal density are spheroids of revolution about the axis of rotation.

On substituting this value of Y_2 into the equation for Y_2, we find

$$-\frac{\epsilon}{r_1}\int_0^{r_1}\rho' a'^2 da' + \frac{1}{5r_1^3}\int_0^{r_1}\rho' \frac{d}{da'}(a'^5\epsilon')\, da' + \frac{r_1^2}{5}\int_{r_1}^{a}\rho' \frac{d\epsilon'}{da'}\, da' = -\frac{1}{8\pi f}\Omega^2 r_1^2 \qquad \text{.........(3).}$$

Let us multiply by r_1^3, and then differentiate with regard to r_1. We find on simplifying

$$-\left(r_1^2 \frac{d\epsilon}{dr_1} + 2\epsilon r_1\right)\int_0^{r_1}\rho' a'^2 da' + r_1^4\int_{r_1}^{a}\rho' \frac{d\epsilon'}{da'}\, da' = -\frac{5}{8\pi f}\Omega^2 r_1^4,$$

or, dividing by r_1^4,

$$-\left(\frac{1}{r_1^2}\frac{d\epsilon}{dr_1} + \frac{2\epsilon}{r_1^3}\right)\int_0^{r_1}\rho' a'^2 da' + \int_{r_1}^{a}\rho' \frac{d\epsilon'}{da'}\, da' = -\frac{5}{8\pi f}\Omega^2 \qquad \text{.........(4).}$$

Differentiating again with regard to r_1, we have

$$\left(\frac{1}{r_1^2}\frac{d^2\epsilon}{dr_1^2} - \frac{6\epsilon}{r_1^4}\right)\int_0^{r_1}\rho' a'^2 da' + 2\left(\frac{d\epsilon}{dr_1} + \frac{\epsilon}{r_1}\right)\rho = 0 \qquad \text{.........(5).}$$

We now introduce the function $S(r_1)$, given by 13·13 (2),

$$S(r_1) = 3\int_0^{r_1} \rho' a'^2 da' \qquad \text{.........(6).}$$

We may at this stage replace r_1 by r without confusion. Then (5) can be written in the forms

$$\frac{d^2}{dr^2}\{S(r)\,\epsilon\} = \frac{6}{r^2}S(r)\,\epsilon + 3r^2\epsilon\frac{d\rho}{dr} \qquad \text{.........(7),}$$

and

$$\frac{d^2\epsilon}{dr^2} + \frac{6\rho r^2}{S(r)}\frac{d\epsilon}{dr} - \left(1 - \frac{\rho r^3}{S(r)}\right)\frac{6\epsilon}{r^2} = 0 \qquad \text{.........(8).}$$

13·5. *Two Theorems of Clairaut*. We are now in a position to prove that the ellipticities of the strata of equal density increase steadily from the centre to the surface. The density must increase steadily inwards, partly because the heaviest materials would sink to the centre in the fluid earth, and partly because the nearer the centre the greater the pressure, and therefore the more the material is compressed. Thus $d\rho/dr$ is essentially negative. Now

$$S(r) = \int_0^r \rho' da'^3 = \rho r^3 - \int_0^r a'^3 \frac{d\rho'}{da'} da' \qquad \text{.........(1),}$$

and the second term of this is always negative. Hence $S(r)$ is always greater than ρr^3, and $1 - \rho r^3/S(r)$ is always positive. When r is small, the latter quantity tends to zero at least as fast as r. Let the first term in it that does not vanish be Hr^k. Then k is at least unity, and H is positive. Now let the expansion of ϵ in powers of r start with r^p. On substituting in 13·4 (8) and equating coefficients of r^{p-2} we see that

$$p(p+5) = 0 \qquad \text{.........(2),}$$

whence $p = 0$ or -5. The negative root is impossible, since it would make ϵ infinite at the centre. Therefore the only admissible root of (2) is zero, and ϵ is finite at the centre, equal to A, say. If the next term in ϵ that does not vanish is Br^s, we may substitute $A + Br^s + \ldots$ for ϵ in 13·4 (8). The terms of lowest degree are $Bs(s+5)\,r^{s-2} - 6Hr^{k-2}A$. Since A and H are not zero, the term in B must cancel the term in AH, and therefore $s = k$; and since H is positive, B has the same sign as A. Now when r is small $\frac{d\epsilon}{dr}$ behaves like sBr^{s-1}, and since s is positive this must have the same sign as B, and therefore as A. Thus $\frac{d\epsilon}{dr}$ must have the same sign as ϵ when r is small.

Now when $d\epsilon/dr$ and ϵ have the same sign, it follows that ϵ increases numerically with r. Thus the ellipticity increases from the centre. It could only cease to increase if $d\epsilon/dr$ became zero. But if $d\epsilon/dr$ is zero, we see from 13·4 (8) that $d^2\epsilon/dr^2$ must have the same sign as ϵ. Thus for slightly greater values of r, $d\epsilon/dr$ will again have the same sign as ϵ, and ϵ will again proceed to increase numerically. Thus ϵ must increase with r right up to the surface.

13·51. It is not, however, enough that ϵ should satisfy the differential equation 13·4 (8). It must also satisfy the first and second integrals 13·4 (3) and (4) whence this equation was derived, otherwise it will not be a solution of the problem. If in 13·4 (3) we put $r_1 = a$, and use 13·13 (1) and 13·3 (8), we get

$$\int_0^a \rho' \frac{\partial}{\partial a'} (a'^5 \epsilon')\, da' = \frac{5a^2 M}{4\pi} (\epsilon_a - \tfrac{1}{2} m) \qquad \text{.........(1).}$$

This is equivalent to 13·3 (9). The left side of (1) would be increased numerically if ϵ were everywhere the same and equal to its surface value ϵ_a. Then $\int_0^a \rho' \frac{d}{da'} a'^5 da'$ would be increased if matter was removed from the interior to the exterior to make the density uniform. Hence on both grounds

$$\left| \int_0^a \rho' \frac{d}{da'} (a'^5 \epsilon')\, da' \right| < |\epsilon_a| \bar{\rho} a^5 \qquad \text{.........(2),}$$

where $\bar{\rho}$ is the mean density. But

$$\bar{\rho} = \frac{3M}{4\pi a^3} \qquad \text{.........(3),}$$

and therefore the left side is numerically less than $3a^2 M\epsilon_a/4\pi$. Hence (1) shows that ϵ_a has the same sign as m, and is therefore positive. But ϵ never changes sign, and therefore the ellipticities of all strata below the surface are positive. In other words, the value of r for any stratum is greatest when θ is equal to $\frac{1}{2}\pi$; that is, all the strata of equal density are oblate.

13·52. We can employ similar methods to show that the expression for r can contain no spherical harmonics except $\frac{1}{3} - \cos^2\theta$. For by a process similar to that used in 13·4 we can show that Y_n must satisfy the differential equation

$$\frac{d^2 Y_n}{dr^2} - n(n+1)\frac{Y_n}{r^2} + \frac{6\rho}{S(r)}\left(r^2 \frac{dY_n}{dr} + rY_n\right) = 0.$$

If when r is small Y_n behaves like r^p, we readily see that

$$p(p+5) - n^2 - n + 6 = 0.$$

In the first place, if $n = 1$, the values of p that satisfy this equation are -1 and -4. The former gives only a displacement of the earth as a whole as a rigid body; the second is impossible. Therefore the only possible first harmonic displacement is irrelevant.

If n is equal to or greater than 2, there is one zero or positive value of p, and this, by an argument exactly like that of 13·5, makes Y_n increase steadily numerically with r. Then as in 13·51 we can show that when r_1 is put equal to a, the first term in 13·4 (1) is $\bar{\rho} a^2 Y_n$, while the third is zero and the second is numerically less than $\frac{1}{2n+1} \bar{\rho} a^2 Y_n$. Thus 13·4 (1) cannot be satisfied unless Y_n is zero. It follows that the harmonic in $\frac{1}{3} - \cos^2\theta$ is

the only one present in the figure of the earth on the theory of hydrostatic equilibrium.

13·53. We can now prove that ϵ/r^3, which is evidently always positive, decreases steadily as r increases. In 13·4 (8) let us put

$$\epsilon = \lambda r^3 \qquad \text{.........(1).}$$

Then we find
$$\frac{d^2\lambda}{dr^2} + 6\left(\frac{\rho r^2}{S(r)} + \frac{1}{r}\right)\frac{d\lambda}{dr} + \frac{24\rho r}{S(r)}\lambda = 0 \qquad \text{.........(2).}$$

Now ϵ is finite when r is zero, and therefore λ behaves like r^{-3} when r is small. It therefore diminishes as r increases. It could cease to diminish only if $d\lambda/dr$ became zero; but then (2) shows that $d^2\lambda/dr^2$ would be negative, and therefore for slightly greater values of r, $d\lambda/dr$ would again be negative, and λ would continue to diminish. Thus ϵ/r^3 decreases steadily as r increases. The theorems of 13·5 and 13·53 are due to Clairaut.

13·54. If, again, we put r equal to a in 13·4 (4), we get

$$\left\{\frac{1}{a^2}\left(\frac{d\epsilon}{dr}\right)_a + \frac{2\epsilon_a}{a^3}\right\}\int_0^a \rho' a'^2 da' = \frac{5}{8\pi f}\Omega^2 \qquad \text{.........(1),}$$

where the suffix a indicates that the corresponding quantity is to be evaluated at the surface. Using now 13·13 (3) and 13·3 (8) we have

$$a\left(\frac{d\epsilon}{dr}\right)_a + 2\epsilon_a = \tfrac{5}{2}m \qquad \text{.........(2).}$$

13·6. Let us now introduce a new dependent variable η, defined by

$$\eta = \frac{d\log\epsilon}{d\log r} = \frac{r}{\epsilon}\frac{d\epsilon}{dr} \qquad \text{.........(1).}$$

Then
$$\frac{d\epsilon}{dr} = \frac{\eta\epsilon}{r};\quad \frac{d^2\epsilon}{dr^2} = \left(\frac{1}{r}\frac{d\eta}{dr} + \frac{\eta^2}{r^2} - \frac{\eta}{r^2}\right)\epsilon \qquad \text{.........(2).}$$

Substituting in 13·4 (8)

$$\frac{rd\eta}{dr} + \eta^2 - \eta + \frac{6\rho r^3}{S(r)}\eta - 6\left(1 - \frac{\rho r^3}{S(r)}\right) = 0 \qquad \text{.........(3).}$$

Let us now introduce the auxiliary function ρ_0, given by

$$\rho_0 = \frac{1}{r^3}\int_0^r \rho\, dr^3 \qquad \text{.........(4),}$$

so that ρ_0 is the mean density of all the matter within distance r of the centre. Then (4) can be transformed to

$$\rho r^2 = \frac{1}{3}\frac{d}{dr}(\rho_0 r^3) \qquad \text{.........(5),}$$

whence
$$\frac{\rho r^3}{S(r)} = 1 + \frac{1}{3}\frac{r}{\rho_0}\frac{d\rho_0}{dr} \qquad \text{.........(6),}$$

and (3) becomes
$$r\frac{d\eta}{dr} + \eta^2 + 5\eta + 2\frac{r}{\rho_0}\frac{d\rho_0}{dr}(1+\eta) = 0 \qquad \text{.........(7).}$$

Now we have the identity, by logarithmic differentiation,

$$\frac{\dfrac{d}{dr}\{\rho_0 r^5\sqrt{(1+\eta)}\}}{\rho_0 r^5\sqrt{(1+\eta)}}=\frac{1}{\rho_0}\frac{d\rho_0}{dr}+\frac{5}{r}+\frac{1}{2(1+\eta)}\frac{d\eta}{dr} \qquad \text{.........(8).}$$

This may be used to eliminate $\dfrac{d\eta}{dr}$ from (7). Then

$$\frac{2\sqrt{(1+\eta)}}{\rho_0 r^4}\frac{d}{dr}\{\rho_0 r^5\sqrt{(1+\eta)}\}=10\,(1+\tfrac{1}{2}\eta-\tfrac{1}{10}\eta^2) \qquad \text{.........(9),}$$

which can be written

$$\frac{d}{dr}\{\rho_0 r^5\sqrt{(1+\eta)}\}=5\rho_0 r^4\,\frac{1+\frac{1}{2}\eta-\frac{1}{10}\eta^2}{\sqrt{(1+\eta)}} \qquad \text{.........(10).}$$

This equation is due to Radau*. Its importance rests on the remarkable properties of the function

$$\psi(\eta)=\frac{1+\frac{1}{2}\eta-\frac{1}{10}\eta^2}{\sqrt{(1+\eta)}} \qquad \text{.........(11).}$$

We have

$$\frac{1}{\psi}\frac{d\psi}{d\eta}=\frac{1}{20}\,\frac{\eta\,(1-3\eta)}{(1+\eta)\,(1+\frac{1}{2}\eta-\frac{1}{10}\eta^2)} \qquad \text{.........(12),}$$

so that ψ has a minimum for $\eta=0$ and a maximum for $\eta=\frac{1}{3}$; for large values it steadily diminishes as η increases. When $\eta=0$, $\psi=1$ exactly; when $\eta=\frac{1}{3}$, $\psi=1{\cdot}00074$. When $\eta=0{\cdot}56$, $\psi=0{\cdot}99929$, and when $\eta=3$, $\psi=0{\cdot}8$.

Now on referring back to the result of 13·52, we can write

$$\frac{r^3}{\epsilon}\frac{d}{dr}\left(\frac{\epsilon}{r^3}\right)>0 \qquad \text{.........(13),}$$

or

$$\frac{3}{r}-\frac{1}{\epsilon}\frac{d\epsilon}{dr}>0 \qquad \text{.........(14).}$$

Hence, by the definition of η, η is essentially less than 3.

Again, 13·53 (2) can be written

$$\begin{aligned}\eta_a&=\frac{5}{2}\frac{m}{\epsilon_a}-2\\&=0{\cdot}569\end{aligned} \qquad \text{.........(15),}$$

by the data of 13·3.

When $r=0$, ϵ is finite, and $d\epsilon/dr$ is not infinite. Hence η is zero. It follows that in the earth, ψ is unity at the centre, rises to 1·00074 at the stratum where $\eta=\frac{1}{3}$, and sinks to 0·99929 at the surface. Except in the very improbable event that η can make a wide excursion beyond the limits it attains at the ends of the range, it follows that ψ can never differ from unity by more than 8 parts in 10,000. Thus with an accuracy of this order

$$\frac{d}{dr}\{\rho_0 r^5\sqrt{(1+\eta)}\}=5\rho_0 r^4 \qquad \text{.........(16).}$$

* *Comptes Rendus*, **100**, 1885, 972–77.

This result has been used by Darwin* to approximate to the moment of inertia of the earth. For

$$C = \tfrac{8}{3}\pi \int_0^a \rho r^4 dr \qquad \text{.........(17)},$$

neglecting small quantities; and on replacing ρ by ρ_0 by means of (5)

$$C = \tfrac{8}{9}\pi \int_0^a \left(3r^4\rho_0 + r^5 \frac{d\rho_0}{dr}\right) dr$$

$$= \tfrac{8}{9}\pi \left[\rho_{0a} a^5 - 2\int_0^a \rho_0 r^4 dr\right] \qquad \text{.........(18)},$$

on integrating the second term by parts. But by integration of (16)

$$\int_0^a \rho_0 r^4 dr = \tfrac{1}{5}\rho_{0a} a^5 \sqrt{(1+\eta_a)} \qquad \text{.........(19)}.$$

Thus $$C = \tfrac{8}{9}\pi\rho_{0a} a^5 \{1 - \tfrac{2}{5}\sqrt{(1+\eta_a)}\} \qquad \text{.........(20)}.$$

But evidently ρ_{0a} is the mean density, so that

$$M = \tfrac{4}{3}\pi\rho_{0a} a^3 \qquad \text{.........(21)}.$$

Thus by division $$\frac{C}{Ma^2} = \tfrac{2}{3}\{1 - \tfrac{2}{5}\sqrt{(1+\eta_a)}\} \qquad \text{.........(22)}.$$

Combining this with 13·3 (9) we get

$$\frac{C-A}{C} = \frac{\epsilon_a - \frac{1}{2}m}{1 - \frac{2}{5}\sqrt{(1+\eta_a)}} \qquad \text{.........(23)}.$$

Now η_a is a known function of ϵ_a and m, by (15). Thus (23) is a relation connecting $(C-A)/C$, ϵ_a, and m, and involving no hypothetical part. Given any two of these, it should therefore be possible to calculate the third. Of the three, ϵ_a is the least accurately known from other sources. Taking then as our data

$$\frac{C-A}{C} = 0{\cdot}003272 \qquad \text{.........(24)},$$

$$m = 0{\cdot}003467 \qquad \text{.........(25)},$$

we may solve by putting $$\epsilon_a = m\,(1-\delta) \qquad \text{.........(26)},$$

giving $$0{\cdot}9436 = \frac{\frac{1}{2} - \delta}{1 - \dfrac{2}{5}\sqrt{\dfrac{5}{2\,(1-\delta)} - 1}} \qquad \text{.........(27)},$$

which can be solved for δ by expanding and neglecting δ^3. We find

$$\delta = 0{\cdot}0319 \qquad \text{.........(28)},$$

whence $$\epsilon = \frac{1}{297{\cdot}9} \qquad \text{.........(29)},$$

agreeing with 13·3 (21) within a quantity of the order of ϵ^2.

In Darwin's monumental paper quantities of the order of the squares of the ellipticity have been retained. He gets

$$\epsilon = \frac{1}{296{\cdot}4} \qquad \text{.........(30)}.$$

* *Scientific Papers*, **3**, 78–118; or *M.N.R.A.S.* **60**, 1900, 82–124.

13·7. *Observed Values of the Ellipticity.* At this stage it will be well to collect the chief estimates of the earth's ellipticity that have hitherto been made. Clarke's famous trigonometrical determination of 1866 gave 1/294·9. Hayford, in 1909, from the trigonometric survey of the United States, obtained the value $1/297{\cdot}0 \pm 0{\cdot}5$. This value was inferred from only a limited region of the earth's surface, and to that extent is imperfect; on the other hand, it is superior to all other trigonometric results in that isostasy was taken into account in reducing the observations. Helmert, from measurements of the intensity of gravity, inferred that the ellipticity was $1/298{\cdot}3 \pm 0{\cdot}7$. Hayford and Bowie*, from gravity measurements in the United States, got $1/298{\cdot}4 \pm 1{\cdot}5$. Helmert†, in 1915, revised his earlier estimate to $1/296{\cdot}7 \pm 0{\cdot}6$. Bowie‡, from a comparison of gravity observations all over the earth, has obtained what is probably the best observational result extant, namely 1/297·4. It is seen that Darwin's theoretical determination from the precessional constant, namely 1/296·4, is a trifle larger than the best modern observational results; but the discrepancy is of the order of magnitude both of the square of the ellipticity and of the probable errors of the experimental determinations. If the difference is real, it must be attributed to the earth's not being in a hydrostatic state, and therefore gives an estimate of the extent of the departure from that state. We may say accordingly that the second harmonic inequality in the figure of the earth is isostatically compensated to within something of the order of 70 metres.

13·8. *Possible Distributions of Internal Density.* We have seen that a knowledge of the values of the precessional constant and m, the ratio of the centrifugal force to gravity at the equator, is enough to fix the surface ellipticity, by 13·6 (23), and that then the surface ellipticity, the precessional constant, and m are together enough to fix the moment of inertia of the earth about its axis, by 13·6 (22). The only opportunity of confronting the theory with fact is provided by the surface ellipticity, and the agreement is satisfactory. It appears, therefore, that the form of the law of density in the interior of the earth is not a matter of importance to the theory of the figure of the earth; it cannot affect the surface ellipticity by more than a few parts in 10,000, when the precessional constant and m are known, as they are. Conversely, if we have a hypothetical distribution of density within the earth that gives the correct moment of inertia, $0{\cdot}334Ma^2$, then 13·6 (22) determines η_a, and from this 13·6 (15) determines m/ϵ_a and hence ϵ_a when m is known. Finally 13·6 (23) determines the precessional constant. Every determination in unique, and therefore so long as a hypothetical law of density gives the correct moment of inertia, it is bound to give the correct ellipticity of the surface, the

* *U.S. Coast and Geodetic Survey Special Publication*, No. 10, 1912.

† *Sitzber. d. k. Preuss. Akad. d. Wiss.* **41**, 1915, 676–85.

‡ Cf. p. 189.

correct precessional constant, and the correct variation of gravity over the surface. These are all the data we have relevant to the second harmonic inequality in the figure of the earth.

13·81. The above theory was developed independently by Darwin and Callandreau, second degree terms being included. It has very much reduced the interest of special hypotheses concerning the distribution of density within the earth. Several such hypotheses have been framed, but now that it is realized that any law of density that gives the moment of inertia correctly will satisfy all the data equally well, and as nearly as observation can test, there is no longer much reason for elaborate discussion of particular laws. In this section, therefore, the chief suggested laws will be merely outlined. Full accounts of those of Laplace and Roche may be found in Tisserand's *Mécanique Céleste*; Wiechert's law is of special interest, and will be treated at somewhat greater length. The first requirement in framing any of these laws was that it should make the ellipticity of a stratum of equal density expressible in finite terms as a function of its mean radius; in other words, that it should make 13·4 (8) integrable in finite terms.

13·82. Laplace's hypothesis was that

$$3r^2\frac{d\rho}{dr} = -q^2 S(r) \qquad \text{.........(1)},$$

where q is a constant. Then 13·4 (7) becomes

$$\frac{d^2}{dr^2}\{S(r)\,\epsilon\} + \left(q^2 - \frac{6}{r^2}\right) S(r)\,\epsilon = 0 \qquad \text{.........(2)}.$$

The solution of (1) that remains finite at the centre is

$$\rho = \frac{Q}{r}\sin qr \qquad \text{.........(3)},$$

where Q is a further constant. Then

$$S(r) = \frac{Q}{q^2}(\sin qr - qr\cos qr) \qquad \text{.........(4)},$$

$$S(r) \propto \left(1 - \frac{3}{q^2r^2}\right)\sin qr + \frac{3}{qr}\cos qr \qquad \text{.........(5)}.$$

When q and Q are chosen so as to make the mass and the moment of inertia correct, the law is found to make the surface density about 2·8 and the density at the centre about 11.

13·83. Roche's law is $\quad \rho = \sigma(1 - kr^2) \qquad$ ....(1).

It makes the equation for the ellipticity soluble in terms of hypergeometric functions. It gives a central density of 10·10.

13·84. Laplace's and Roche's distributions of density are not physical laws in any sense. They rest on no known data about the constitution

and properties of the matter within the earth. Indeed the little knowledge we have suggests that the interior is so dense as to be probably metallic, while the outside is rocky. Such materials cannot mix freely, and therefore we should expect a fairly sharp boundary between them, with a sudden discontinuity in density. Laplace's and Roche's formulae, however, make the density distribution continuous at all levels. It is necessary to emphasize this point, for each of these formulae, especially Laplace's, is often mentioned as if it corresponded to a true physical law describing the compressibility of the interior of the earth, which is not the case; the only reason for any interest in either formula is that it makes equation 13·4 (8) integrable in finite terms.

13·85. A distribution of density of greater physical interest is that of Wiechert*. In this the earth is supposed composed of two strata of uniform density, the inner and denser material extending from the centre most of the way to the surface, and the lighter occupying the outer regions. There is a marked difference between the densities of the two layers.

Evidently the hypothesis of complete homogeneity in either layer is too extreme. The inner parts of either layer are under greater pressure than the outer parts of the same layer, and therefore must have been compressed more. But if that is so, the inner parts must have been less dense than the outer parts when under no pressure; in other words, the lightest materials of each layer are supposed to have sunk to the bottom. Thus the assumption of even local homogeneity of density on the earth is unjustified. It appears probable, however, that this hypothesis is more accurate than any involving a continuously varying density yet advanced. It possesses the further advantage that it makes the equations of elastic strain in the earth as a whole integrable.

In discussing this hypothesis a slight change of notation will be convenient. Let a and a_1 be the outer radius and the radius of the denser core respectively, and put

$$a_1 = a\alpha \qquad \text{.........(1).}$$

Let the densities of the shell and the core be ρ_0 and ρ_1, and put

$$\rho_1 = \rho_0 (1 + \mu) \qquad \text{.........(2).}$$

Then the mass is given by

$$M = \tfrac{4}{3}\pi\rho_0 a^3 (1 + \mu\alpha^3) \qquad \text{.........(3),}$$

and the moment of inertia by

$$C = \tfrac{8}{15}\pi\rho_0 a^5 (1 + \mu\alpha^5) \qquad \text{.........(4).}$$

Thus our equation 13·3 (25) gives

$$\frac{2}{5}\frac{1 + \mu\alpha^5}{1 + \mu\alpha^3} = 0{\cdot}334 \qquad \text{.........(5),}$$

and, by what has been said in 13·8, if we can find values of α and μ that

* *Gott. Nach.* 1897, 221–43.

will satisfy this equation, all the other data will be satisfied automatically. Now (5) can be transformed into

$$\frac{1}{\mu} = 5\alpha^3 - 6\alpha^5 \qquad \text{.......(6).}$$

By our physical conditions μ must be positive and finite. Hence in order that there may be any solution $6\alpha^2$ must be less than 5. Thus α must be less than 0·913. The corresponding value of ρ_1, which is the least possible for a given mean density, is 1·31 times that density.

In Wiechert's actual solution, α was equal to 0·779. This, with the above data, makes μ equal to 1·63. If now we take the mean density of the earth equal to 5·53, following Boys and Braun, we find

$$\rho_0 = 3{\cdot}12, \quad \rho_1 = 8{\cdot}22.$$

Wiechert's results were $\rho_0 = 3{\cdot}200$, $\rho_1 = 8{\cdot}206$, calculated on the supposition of a mean density of 5·58. If a factor is applied to correct the mean density, his results become $\rho_0 = 3{\cdot}17$, $\rho_1 = 8{\cdot}15$. The discrepancies are probably attributable partly to differences in the numerical data used and partly to the fact that Wiechert included terms of the second degree in the ellipticity in his analysis. The total number of possible solutions is evidently infinite.

The value adopted for ρ_0, the density of the outer shell, corresponds roughly with those of rather dense rocks, and that for the core with those of many common metals, notably iron, copper, and nickel. It is therefore reasonable to suppose that the outer crust is rocky down to some depth of the order of 1300 to 1900 km., and that the materials of the earth below that level are mainly composed of a metallic alloy, probably a nickel-iron alloy. It is interesting to observe that the transition indicated by Wiechert occurs at a depth nearly equal to that where the velocity curves of seismic waves, as determined by Knott, show sudden bends. If this identification is correct, it would imply that in the outer layer the velocities increase with depth, but that in the metallic core they are practically independent of depth. The absence of any discontinuity in velocity at the layer of contact is curious, for it indicates an agreement that could hardly have been predicted between the velocities of waves in different substances. At ordinary pressures the velocity of distortional waves in iron is readily found, from the data in Kaye and Laby's Tables, to be about 3·2 km./sec., which is in good agreement with the velocity in the granitic layer inferred from the Oppau explosion; but it is surprising that a still closer agreement should be found at a great depth.

13·9. *Figure of the Moon.* Let us now proceed to consider the second harmonic inequalities in the figure of the moon. The moon's rotation about its axis must produce an oblateness of its figure; but in addition the attraction of the earth tends to raise two tidal protuberances, one just under the earth, and the other just opposite to it. Both phenomena need

to be considered in any account of the figure of the moon. It will be sufficient in this discussion to treat the moon as homogeneous.

Let axes of x, y, and z be taken at the centre of the moon and moving with it. The x axis points away from the earth, the z axis is perpendicular to the plane of the orbit, and the y axis is perpendicular to both of these. If M be the mass of the earth, m that of the moon, and n the mean angular velocity of the moon about the earth, the chief part of the potential due to the earth's attraction is

$$U = fM/R \qquad \text{.........(1)},$$

where R is the distance from the point considered to the centre of the earth. If c be the moon's mean distance from the earth, we have

$$R^2 = (c + x)^2 + y^2 + z^2 \qquad \text{.........(2)},$$

and therefore
$$U = \frac{fM}{c} - \frac{fMx}{c^2} + \frac{fM}{2c^3}(2x^2 - y^2 - z^2) \qquad \text{.........(3)},$$

to the second order in the radius of the moon. Now we are supposing the motion to be one of steady revolution, the moon always keeping the same face towards the earth, so that the motion of each part of the moon is one of revolution with angular velocity n about an axis perpendicular to the plane of the orbit through the centre of gravity of the earth and moon together. The distance of the latter point from the centre of the moon is $Mc/(M + m)$, or $c/(1 + \mu)$, where

$$\mu = m/M \qquad \text{.........(4)}.$$

The perpendicular distance of a point from the axis of the general rotation is therefore $\left\{\left(\frac{c}{1+\mu} + x\right)^2 + y^2\right\}^{\frac{1}{2}}$. Hence by 13·2 the effect of the steady rotation and revolution in distorting the moon is equivalent to that of a potential

$$\tfrac{1}{2}n^2\left\{\left(\frac{c}{1+\mu} + x\right)^2 + y^2\right\} \qquad \text{.........(5)}.$$

Combining this with (3), we find that the total effective disturbing potential is

$$\frac{fM}{c} - \frac{fM}{c^2}x + \frac{fM}{2c^3}(2x^2 - y^2 - z^2) + \tfrac{1}{2}n^2\frac{c^2}{(1+\mu)^2} + \frac{n^2c}{1+\mu}x + \tfrac{1}{2}n^2(x^2 + y^2) \qquad \text{.........(6)},$$

and then, since
$$n^2c^3 = f(M + m)$$

the two terms in x cancel, and we have altogether a disturbing potential

$$\frac{fM}{2c^3}(2x^2 - y^2 - z^2) + \tfrac{1}{2}n^2(x^2 + y^2)$$

due to the earth and rotation together. We require to know what effect this will have on the figure of the moon.

In the first place, it is clear that the superposition of a small disturbing potential, symmetrical about the centre of the moon, will not affect its ellipticities to the first order of small quantities. Let us include, therefore, a

potential $\lambda(x^2+y^2+z^2)$, λ being afterwards determined so as to make the whole potential a solid harmonic. The requisite condition for this is that

$$n^2 + 3\lambda = 0.$$

The disturbing potential is therefore equivalent to

$$\frac{1}{2}\frac{n^2}{1+\mu}(2x^2-y^2-z^2)+\tfrac{1}{6}n^2(x^2+y^2-2z^2).$$

If we now omit μ, which is about 1/82, this reduces to

$$\tfrac{1}{6}n^2(7x^2-2y^2-5z^2) \qquad\qquad \text{.........(7)},$$

which we may write simply Kr^2S_2.

Now suppose the surface of the moon to become deformed until its equation is

$$r = a(1+\epsilon S_2) \qquad\qquad \text{.........(8)}.$$

We see from 13·11 (2) that its external potential becomes

$$fm\left(\frac{1}{r}+\frac{3}{5}\frac{a^2\epsilon S_2}{r^3}\right) \qquad\qquad \text{.........(9)}.$$

The condition that the moon should be in hydrostatic equilibrium is that the sum of its own external potential and the disturbing potential shall be constant over its surface. Hence $fm\left(\frac{1}{r}+\frac{3}{5}\frac{a^2\epsilon S_2}{r^3}\right)+Kr^2S_2$ must reduce to a constant when (8) is satisfied. This gives

$$\epsilon = \frac{5}{2}K\frac{a^3}{fm} \qquad\qquad \text{.........(10)},$$

and the equation of the moon's surface is

$$\begin{aligned} r &= a\left\{1+\frac{5}{12}n^2\frac{a}{fm}(7x^2-2y^2-5z^2)\right\} \\ &= a\left\{1+\frac{5}{12}\frac{M}{m}\frac{a^3}{c^3}\frac{7x^2-2y^2-5z^2}{a^2}\right\} \qquad\qquad \text{.........(11)}. \end{aligned}$$

Thus the semiaxes of x, y, z on the moon are respectively

$$a\left(1+\frac{35}{12}\frac{M}{m}\frac{a^3}{c^3}\right),\quad a\left(1-\frac{10}{12}\frac{M}{m}\frac{a^3}{c^3}\right),\quad a\left(1-\frac{25}{12}\frac{M}{m}\frac{a^3}{c^3}\right) \quad \text{...(12)}.$$

If then A, B, C be the principal moments of inertia of the moon about its centre, we have

$$\frac{C-A}{C} = 5\frac{M}{m}\frac{a^3}{c^3} = 0{\cdot}0000375 \qquad\qquad \text{.........(13)},$$

$$\frac{C-B}{C} = \frac{5}{4}\frac{M}{m}\frac{a^3}{c^3} = 0{\cdot}0000094 \qquad\qquad \text{.........(14)},$$

where the numerical evaluations are based on current knowledge of the moon's mass and size. In addition we have identically $\frac{B-A}{C-A}=\frac{3}{4}$.

13·91. Now the two ratios $(C-A)/C$ and $(B-A)/C$ for the moon are capable of being found from observation. The former is proportional

to the mean inclination of the moon's equator to the ecliptic, and $(B-A)/C$ to the amplitude of the moon's true libration in longitude. Both these propositions rest on purely dynamical considerations*. The inclination can be determined with much certainty; there seems little room for doubt that the corresponding value of $(C-A)/C$ is close to 0·0006289. The amplitude of the libration in longitude, however, has been the subject of much discussion. The balance of opinion at present seems to follow Hayn† in making the ratio $(B-A)/(C-A)$ nearly equal to its theoretical value 0·75; but several investigators have obtained values in the neighbourhood of 0·5, and the difference appears to be attributable only to observational errors‡. In either case both $(C-A)/C$ and $(C-B)/C$ very much exceed their theoretical values.

The discrepancy was first noticed by Laplace, who was content to attribute the high values of these quantities to distortions developed during solidification. There is no reason on this theory why they should have attained any particular magnitudes; the actual values must be regarded as accidental. Let us see whether it is possible to reconcile the data on the basis of known laws.

In the first place, the numerical coefficients depend on the hypothesis that the moon is homogeneous. If we suppose the moon to be composed of two constituents of the same densities as Wiechert's two layers in the earth, the value of the radius of the inner sphere being determined to fit the actual mean density of the moon, it is found that little change is made. The coefficients are multiplied by 0·9; there is no other change, and the discrepancy is slightly increased.

The present determinations of the masses of the earth and moon, and of the mean radius of the moon, are incapable of serious error, and their values in the past cannot have varied much, at least since the moon solidified.

There remains the mean distance of the moon from the earth. Let us suppose, then, that the moon last adjusted itself to the hydrostatic form when considerably nearer the earth than it is now, at a distance c_1, say, and consider the moon's subsequent development as it receded from the earth and the disturbing influences correspondingly diminished. Since by hypothesis, no further adjustment by plasticity is taking place, the only change of form is that due to pure elastic deformation, and is much less than the change that would take place in a fluid body. Let β_0 denote the value that $(C-A)/C$ would sink to if the moon receded to an infinite distance from the earth without any hydrostatic adjustment taking place. At any time later than the last adjustment, $(C-A)/C$ is composed of

* Routh, *Advanced Rigid Dynamics*, 1905, Chap. 12; Tisserand, *Mécanique Céleste*, t. 2, Chap. 28.

† *Abhand. d. k. Sächs. Gesell. d. Wiss. zu Leipzig*, **30**, **1907**, **1–103**.

‡ Stratton, *Memoirs of R.A.S.* **59**, 1909, 257–90.

two parts, β_0 and the part due to elastic strain. Suppose the latter to be κ times the value of $(C - A)/C$ for a fluid moon at the same distance. Then the conditions at the last hydrostatic adjustment were such that β_0, and the elastic value of $(C - A)/C$ appropriate to distance c_1, together made up the hydrostatic value for the same distance. Now, allowing for the influence of heterogeneity, we have for the hydrostatic value,

$$\frac{C-A}{C} = 4{\cdot}5\frac{M}{m}\frac{a^3}{c^3} \qquad \text{.........(1)},$$

and our condition is
$$\beta_0 + 4{\cdot}5\kappa\frac{M}{m}\frac{a^3}{c_1^3} = 4{\cdot}5\frac{M}{m}\frac{a^3}{c_1^3} \qquad \text{.........(2)}.$$

We see next that the elastic deformation is produced by the contemporary influence of the earth's gravity and the moon's rotation together, and therefore the protuberance always points *exactly* to the earth. Thus the attraction of the earth on this protuberance passes exactly through the centre of the moon, cannot affect the moon's rotation, and thus does not take part in determining the inclination of the moon's axis of rotation. The latter is determined by β_0 alone, and therefore the value of $(C - A)/C$ determined from observations is simply β_0. Thus (2) can be written

$$4{\cdot}5\,(1-\kappa)\frac{M}{m}\frac{a^3}{c_1^3} = 0{\cdot}0006289 \qquad \text{.........(3)}.$$

Now κ is small, and can be shown theoretically to be about 0·013 Also

$$\frac{M}{m} = 81{\cdot}2,$$

$$\frac{a}{c_0} = \frac{1}{220},$$

where c_0 is the moon's present distance. Hence

$$\frac{c_1}{c_0} = 0{\cdot}376.$$

The data can therefore be reconciled by supposing that the moon last adjusted itself to the hydrostatic state when its distance from the earth was about 140,000 km., and its period of revolution $27{\cdot}3\,(0{\cdot}376)^{\frac{3}{2}} = 6{\cdot}3$ of our present days.

It will be seen that the arguments about the variation in $(C - A)/C$ since the last hydrostatic adjustment apply equally well to $(B - A)/C$; therefore the inference that the ratio of these quantities should be 0·75 remains unaltered. If, however, the moon during solidification was executing a libration in longitude of amplitude 40°, this would no longer hold, and the ratio would be reduced to 0·50*.

13·92. The theory just developed requires that the matter within the moon should now depart appreciably from the hydrostatic state, and should have maintained this departure for a long time geologically. It must there-

* Jeffreys, *Mem. R.A.S.* **60**, 1915, 187–217.

fore have a finite strength. The amount of this strength can be evaluated roughly, by using a result of Sir G. H. Darwin*. The greatest stress-difference in a homogeneous spheroid due to a second order zonal inequality over the surface is at the centre, and its magnitude is $\frac{3}{9}\frac{2}{5}g\rho ae$, where g is surface gravity, ρ the density, a the radius of the spheroid, and e its ellipticity. For the moon, with our adopted value of β_0, which would be equal to e if the moon were symmetrical, this would make the stress-difference equal to 2×10^7 dynes/cm.^{2} Since the difference between the moments B and C of the moon is only a quarter of $C - A$, a theory regarding the moon as affected only by a zonal harmonic about its x axis will not be far wrong.

Now the moon, like the earth, was probably once fluid, and has been cooling from the surface for about the same time. We should therefore expect that cooling, and with it strength, would have developed to much the same depth in both. Now it was seen in our discussion of isostasy that there is in the earth, between 100 km. and 400 km. in depth, an extensive region where the strength lies between 3×10^7 and 10^8 dynes/cm.^{2} Such a region in the moon, whose radius is only 1700 km., would extend a considerable fraction of the way to the centre, occupying more than half the volume. Its strength, again, is decidedly more than has just been seen to be necessary to enable a homogeneous body to maintain the lunar inequalities of figure. It is therefore probable that the strength would be sufficient to prevent hydrostatic adjustment, even if the matter near the centre has, as is probable, no appreciable permanent strength. If, however, there is a weak region at the centre of the moon, it will have the effect of increasing the elastic strain under a given field of force, and the estimate already made of κ will have to be somewhat increased. It is unlikely, however, that the increase will be great enough to affect the inferred value of c_1 by more than a few per cent.

* *Scientific Papers*, 2, 474–81.

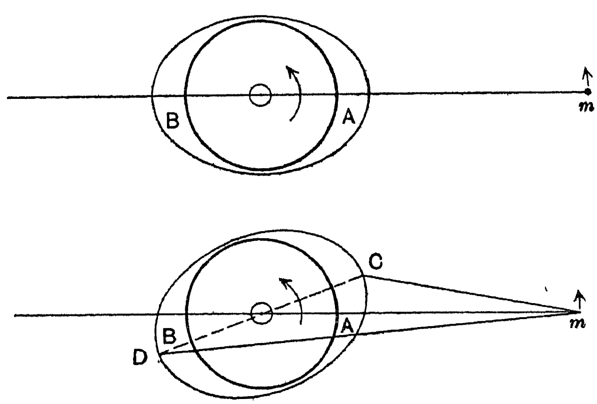
CHAPTER XIV

Tidal Friction

"Ye fyul, sez a chep, it's a bonny myun,
They've ketched and myed it a clock fyece."
Tyneside song, 'The Fiery Clock Fyece.'

14·1. *Nature of Tidal Friction.* In our discussion of the resonance theory of the origin of the moon, it was seen that this theory implies that the moon was initially within a few thousand kilometres of the earth's surface, and that the earth at that time rotated in a few hours. In these circumstances the angular momentum of the earth's rotation much exceeded that attributable to the moon's revolution. On the other hand, the angular momentum of the moon's revolution is now about five times that of the earth's rotation, and it was assumed as essential to the resonance theory that some physical process exists that has been competent to produce the requisite reduction in the speed of rotation of the earth and the associated increase in the mean distance of the moon.

Tidal friction is qualitatively, and probably quantitatively, capable of producing such an effect. The nature of the action of the tides in slowing up the earth's rotation and driving the moon further away may be illustrated as follows. Let m in Figs. 7 and 8 denote the moon, while the circle enclosed by the thick line represents the equatorial section of the solid earth. The arrows indicate the directions of rotation and revolution.



Figs. 7 and 8.

If the solid earth were a perfect sphere, devoid of friction, the tides raised by the moon in a deep ocean would be at A and B, vertically below the moon and straight opposite to it, as shown by the continuous ellipse. Moderate rotation would affect the height of the tides, but not their position with reference to the moon. During the revolution of a particular point P on the

earth's surface, the water level above it will still rise and fall when rotation is taken into account, the maximum tide height occurring when the point is just below or just opposite to the moon, and the minima at the two positions of quadrature. If, however, friction is present, it will delay the times of maximum and minimum elevation at a particular station, in accordance with the well-known effect of friction in making small oscillations lag in phase. Thus the highest tide at P will not occur till some time after it has passed A or B, and the form of the section of the water level by the equator will resemble the ellipse in Fig. 8, the high tides being at C and D.

Now let us consider the attraction of the tides on the moon. For simplicity the two high tides may be replaced by two heavy particles at the opposite points C and D. The attraction of C on the moon is along mC, and that of D is along mD. That of C is the greater, for it is nearer to m. Neither force acts accurately along the radius joining the centre of the earth to the moon, and therefore both have components at right angles to it. That arising from C is evidently the greater, both because the resultant force due to C is the greater, and because the angle CmO is greater than DmO. Thus there will be on the whole a force on the moon with a component in the direction of its revolution.

Similarly we may consider the attraction of the moon on the two tidal protuberances. We see that the force on C is greater than that on D, and further, on account of the fact that the angle CmO is greater than DmO, its line of action passes further from the centre of the earth. On both grounds it has a greater moment about the centre of the earth. Therefore an effect of the moon's attraction on the tides is to produce a couple tending to turn the earth in the direction opposite to its rotation, and thus to slow up its rotation.

Similar results will evidently hold even if there is no ocean, but the interior of the earth is imperfectly elastic. On account of its elasticity, the form of the solid earth will be somewhat distorted by the moon's tidal action; and imperfect elasticity will cause the greatest elevations to be, not at A and B, but at places like C and D. Thus bodily tides, like oceanic tides, will tend to accelerate the moon and retard the earth.

14·2. *Effects of Tidal Friction.* The above simple discussion does not consider the complications introduced by the irregular form of the oceans. It is not to be expected, however, that such a complication will prevent the existence of the couples. The solar tides, further, behave similarly to those raised by the moon, and therefore have to be taken into account in a quantitative discussion of the effects of tidal friction.

Let the masses of the earth, moon and sun be respectively M, m, and m_1. Let the mean angular velocities of the moon and sun about the earth be n and n_1, and their distances from the earth c and c_1. Let the retarding

couples acting on the earth due to the lunar and solar tides be respectively $-N$ and $-N_1$. Let the earth's angular velocity of rotation be Ω, and its principal moment of inertia about its polar axis C. Then

$$n^2c^3 = f(M + m);\ \ n_1{}^2c_1{}^3 = f(m_1 + M) \qquad \text{.........(1)},$$

where f is the constant of gravitation. Put

$$c = c_0\xi^2,\ n = n_0\xi^{-3};\ c_1 = c_{10}\xi_1{}^2,\ n_1 = n_{10}\xi_1{}^{-3} \qquad \text{.........(2)},$$

where the zero suffixes indicate the present values of the quantities they are attached to.

The moon revolves about the centre of gravity of the earth and moon together, the distance of which from the moon is $Mc/(M+m)$. The angular velocity of the moon being n, the linear velocity is $Mcn/(M+m)$, and therefore the angular momentum of the moon about the centre of the earth is $Mc^2n/(M+m)$. That of the earth's rotation is $C\Omega$. By Newton's Third Law, to the couple $-N$ acting on the earth must correspond a force acting on the moon, whose moment about the centre of the earth is $+N$. The solar tides will have no secular effect on the moon, for their period is different, and therefore in the course of a lunar synodic month they are presented to the moon in all aspects. Thus they will accelerate the moon in one half of the month as much as they retard it in the other half. Thus the solar tides will not affect the moon; similarly the lunar tides, in the long run, will not affect the sun. Now noticing that $c^2n = c_0{}^2n_0\xi$, we see by taking moments about the centre of the earth for the moon, sun, and earth respectively that

$$\frac{Mm}{M+m}c_0{}^2n_0\frac{d\xi}{dt} = N \qquad \text{.........(3)},$$

$$\frac{m_1M}{m_1+M}c_{10}{}^2n_{10}\frac{d\xi_1}{dt} = N_1 \qquad \text{.........(4)},$$

$$C\frac{d\Omega}{dt} = -N - N_1 \qquad \text{.........(5)}.$$

If E be the total mechanical energy in the system, it decreases at a rate equal to the sum of the rates of performance of work by the angular motions in overcoming the operating couples. This gives

$$-\frac{dE}{dt} = (N + N_1)\Omega - Nn - N_1n_1 \qquad \text{.........(6)}.$$

Consider the effects of the lunar tides separately. These give

$$-\frac{dE}{dt} = N(\Omega - n) \qquad \text{.........(7)}.$$

Since the couples arise from dissipation of energy, the left side is essentially positive. Thus N has the same sign as $\Omega - n$. In the earth-moon system this is positive, and therefore N is positive. This argument is independent of any assumption about the form of the ocean or the nature of the elasticity of the earth. Similarly we can see that N_1 must be positive. It follows from (3) and (4) that the mean distances of the moon and sun must

be increasing, and accordingly their mean motions, in comparison with an inertial system, must be decreasing. It follows again from (5) that the rate of rotation of the earth must be decreasing.

14·21. Consider now what the ratio of N to N_1 is likely to be. The potential at the surface of a body due to a mass m at distance c can be put in the form

$$\frac{fm}{R} = \frac{fm}{c}\left(1 - \frac{x}{c} + \frac{2x^2 - y^2 - z^2}{2c^2}\right) \qquad \text{.........(1)},$$

where the axis of x joins the centres of the two bodies, that of z is the polar axis, and that of y is perpendicular to both. The acceleration of the body as a whole has components

$$\left(-\frac{fm}{c^2}, 0, 0\right) = -\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\frac{fmx}{c^2} \qquad \text{.........(2)}.$$

The distortion of the body is due to the difference between the actual potential and the potential that would displace it without changing its form. Apart from a constant term, this difference is equal to

$$\frac{fm}{2c^3}(2x^2 - y^2 - z^2) \qquad \text{.........(3)}.$$

We may now introduce spherical polar coordinates at the centre of the earth, so that

$$\left.\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned}\right\} \qquad \text{.........(4)}.$$

Then θ is the colatitude of the point considered, and ϕ is the difference between its longitude and that of the moon. (3) becomes

$$\frac{fmr^2}{2c^3}\{(\tfrac{1}{2} - \tfrac{3}{2}\cos^2\theta) + \tfrac{3}{2}\sin^2\theta\cos 2\phi\} \qquad \text{.........(5)}.$$

The two terms of this are spherical harmonics. The first is independent of the longitude of the point considered, and therefore can produce only a permanent tide not moving with the moon. The second term generates the semidiurnal tide, for we see that it has maxima where $\phi = 0$ and $\phi = \pi$; that is, just below and just opposite to the moon. Thus the tide-generating potential is

$$U = \frac{3}{4}\frac{fmr^2}{c^3}\sin^2\theta\cos 2\phi \qquad \text{.........(6)}.$$

The tide raised by the moon will be expressible as the resultant of a number of surface harmonic displacements. The couple due to the attraction of the moon on an element of the earth's mass is $\rho\dfrac{\partial U}{\partial \phi}d\tau$, where $d\tau$ is an element of volume and ρ the local density. The total couple is therefore

$$-N = -\iiint \frac{3}{2}\frac{fmr^2\rho}{c^3}\sin^2\theta\sin 2\phi\, r^2 dr\, d\omega \qquad \text{.........(7)},$$

where $d\omega$ is an elementary solid angle, so that

$$d\tau = r^2 dr\, d\omega \qquad \text{.........(8)}.$$

The integral is to be taken throughout the earth. If we consider only the oceanic tide, ρ is a function of r and θ only within the solid earth, and therefore the integral up to the mean position of the ocean surface contributes nothing to the couple. If the height of the oceanic tide is h, and the radius of the earth is A, the integral up to the actual ocean surface becomes

$$-N = -\iint \frac{3fmA^4}{2c^3} \rho h \sin^2\theta \sin 2\phi \, d\omega \qquad \text{.........(9)},$$

where ρ now refers to the ocean alone. We see that none of the harmonics involved in h, save that in $\sin^2\theta \sin 2\phi$, can contribute anything to the couple. Let this one be expressible in the form $H \sin^2\theta \cos 2(\phi - \epsilon)$. Evidently 2ϵ is the phase lag introduced by tidal friction, corresponding to double the angle COA in Fig. 8. The portion of this harmonic involving $\cos 2\phi \cos 2\epsilon$ vanishes on integration with regard to ϕ, and we are left with

$$\begin{aligned} -N &= -\frac{3fmA^4}{2c^3} \rho H \int_0^\pi \int_0^{2\pi} \sin^5\theta \sin^2 2\phi \sin 2\epsilon \, d\theta \, d\phi \\ &= -\frac{8}{5}\pi \frac{fmA^4}{c^3} \rho H \sin 2\epsilon \qquad \text{.........(10)}. \end{aligned}$$

A similar discussion is applicable to the solar tides.

If suffix 1 indicates that the quantity it is attached to refers to the sun or to the solar tide, we evidently have from (10)

$$N/N_1 = \frac{mH}{c^3} \sin 2\epsilon \Big/ \frac{m_1 H_1}{c_1^3} \sin 2\epsilon_1 \qquad \text{.........(11)}.$$

We notice that the ratio $\dfrac{m}{c^3}\Big/\dfrac{m_1}{c_1^3}$ is the ratio of the amplitudes of the two tide-generating potentials due to the moon and sun. If then the periods of the two tides were equal, and if the laws determining the tides were linear in the displacements, we could calculate the two tides separately; their phase lags would then be equal, and their amplitudes would be in the ratio of those of the disturbing potentials. Thus we should have

$$\frac{N}{N_1} = \left(\frac{m/c^3}{m_1/c_1^3}\right)^2 \qquad \text{.........(12)}.$$

This ratio is about 5·1.

14·22. We readily see that the above argument can be generalized so as to be applicable to the bodily tide. The displacements at all points of the interior will be nearly in proportion in the lunar and solar tides, provided only that the periods are far from resonance. This hypothesis is justified by the fact that the period of free vibration of a fluid earth would be less than two hours, and this would be shortened by elasticity; while the periods of both tides are about 12 hours. Thus internal motion will introduce into (9) and (10) only such factors as will be the same for both moon and sun, and therefore (11) will hold. But the conditions for inferring (12) from (11) are also satisfied, for we cannot suppose that the

small strains involved in bodily tides produce friction proportional to any power but the first. Thus in this case

$$N/N_1 = 5{\cdot}1.$$

14·23. Lags in the ocean tides produced by viscosity give the same ratio. Much tidal friction, however, takes the form of skin friction in regions of strong tidal currents; it will indeed be seen later that this is probably the most important type operative at the present time. Skin friction is proportional to the square of the velocity of the tidal current, and therefore it is no longer justifiable to treat the lunar and solar tides separately and add the results. If, however, friction is too small to alter seriously the character of the tides, we may estimate the ratio in another way. The effect of friction is to introduce a reaction from the sea where it occurs upon the water of the ocean as a whole. Friction changes sign with the tidal current, and the reaction will therefore contain a portion with a period equal to that of the tide; this portion will produce a small tide in the ocean, out of phase with the tide unaffected by friction. The tide unaffected by friction gives by hypothesis no dissipation of energy and therefore no tidal couples; these arise from the attraction of the moon and sun on the small tides produced by the reaction caused by friction.

In a region where friction is considered, let u be the velocity of the lunar current, and v that of the solar current. Suppose the origin of time so chosen that we can write

$$u = a \sin pt; \quad v = b \sin qt \qquad \text{.........(1).}$$

Evidently b is considerably less than a, and their ratio is approximately the same all over the earth. Further,

$$p = 2(\Omega - n); \quad q = 2(\Omega - n_1) \qquad \text{.........(2).}$$

To a first approximation only we may suppose the whole velocity to be $a \sin pt + b \sin qt$. We require to use this so as to proceed to an approximation to the form of the frictional forces. The frictional force is $F = k(u + v)^2$, directed against the resultant velocity; or, writing it so as to put in evidence the fact that it changes sign with the velocity,

$$F = k \mid u + v \mid (u + v) \qquad \text{.........(3),}$$

where $\mid u + v \mid$ denotes the absolute value of $u + v$.

Let the two parts of F whose periods are equal to those of the semi-diurnal tides be $P \sin pt$ and $Q \sin qt$. Other parts can give only tides whose periods are different from those of the disturbing potentials, and will therefore give only periodic effects on the moon and sun. We need therefore consider only these two terms. Now

$$\pi P = \lim_{n\to\infty} \frac{1}{n}\int_0^{2n\pi} F \sin pt\, dpt \qquad \text{.........(4),}$$

$$\pi Q = \lim_{n\to\infty} \frac{1}{n}\int_0^{2n\pi} F \sin qt\, dqt$$

$$= \frac{q}{p} \lim_{n\to\infty} \frac{1}{n}\int_0^{2n\pi} F \sin qt\, dpt \qquad \text{.........(5).}$$

Put
$$pt = \theta,\ q/p = r \qquad \text{.........(6)}.$$
Then
$$\pi P = \operatorname{Lim} \frac{1}{n} \int_0^{2n\pi} k \mid a \sin\theta + b \sin r\theta \mid (a \sin\theta + b \sin r\theta) \sin\theta \, d\theta \quad \text{...(7)},$$
$$\pi Q = \operatorname{Lim} \frac{q}{pn} \int_0^{2n\pi} k \mid a \sin\theta + b \sin r\theta \mid (a \sin\theta + b \sin r\theta) \sin r\theta \, d\theta \quad \text{...(8)}.$$
Neglecting b in (7) we find
$$\pi P = 2ka^2 \int_0^{\pi} \sin^3\theta \, d\theta = \tfrac{8}{3} ka^2 \qquad \text{.........(9)}.$$
Let θ_m be the mth zero of $(a \sin\theta + b \sin r\theta)$, not counting the origin. If we neglect b^2, we see easily that
$$\theta_m = m\pi - (-1)^m \frac{b}{a} \sin rm\pi \qquad \text{.........(10)},$$
$$\pi Q = \operatorname{Lim} \frac{kq}{pn} \sum_{m=0}^{2n-1} (-1)^m \int_{\theta_m}^{\theta_{m+1}} (a \sin\theta + b \sin r\theta)^2 \sin r\theta \, d\theta \quad \text{......(11)}.$$
The general term in the sum in (11) is
$$\tfrac{1}{2}(-1)^{m-1} \left[a^2 \left\{ \frac{1}{r} \cos r\theta - \frac{1}{2(2+r)} \cos(2+r)\theta + \frac{1}{2(2-r)} \cos(2-r)\theta \right\} \right.$$
$$\left. + 2ab \left\{ \cos\theta - \frac{1}{2(2r+1)} \cos(2r+1)\theta + \frac{1}{2(2r-1)} \cos(2r-1)\theta \right\} + \text{terms in } b^2 \right]_{\theta_m}^{\theta_{m+1}}$$
$$= \tfrac{1}{2}(-1)^{m-1} \left[a^2 \frac{1}{r} \{\cos r(m+1)\pi - \cos rm\pi\} \right.$$
$$- \frac{a^2}{2(2+r)} \{\cos(2+r)(m+1)\pi - \cos(2+r)m\pi\}$$
$$\left. + \frac{a^2}{2(2-r)} \{\cos(2-r)(m+1)\pi - \cos(2-r)m\pi\} \right]$$
$$+ \tfrac{1}{4} ab \{\cos 2r(m+1)\pi - \cos 2rm\pi\}$$
$$- \tfrac{1}{8} ab \{\cos 2(1+r)(m+1)\pi - \cos 2(1+r)m\pi\}$$
$$- \tfrac{1}{8} ab \{\cos 2(1-r)(m+1)\pi - \cos 2(1-r)m\pi\}$$
$$+ 2ab \left[1 - \frac{(-1)^{m-1}}{4(2r+1)} \{\cos(2r+1)(m+1)\pi - \cos(2r+1)m\pi\} \right.$$
$$\left. + \frac{(-1)^{m-1}}{4(2r-1)} \{\cos(2r-1)(m+1)\pi - \cos(2r-1)m\pi\} \right] \qquad \text{.........(12)}.$$
These terms are to be added for all values of m from zero to $2n-1$. Evidently the terms with m in the argument form several series of cosines of angles in arithmetical progression, and therefore their sum never exceeds a definite finite amount. Thus the integral gives
$$\pi Q = \operatorname*{Lim}_{n \to \infty} \frac{kq}{pn} 2ab(2n + \lambda),$$
where λ does not tend to infinity with n
$$= 4kqab/p \qquad \text{.........(13)}.$$

Thus
$$\frac{Q}{P}=\frac{3}{2}\frac{q}{p}\frac{b}{a} \qquad \text{.........(14).}$$

This gives us the required approximation to the ratio of the reactions. Now since by hypothesis the ocean is far from resonance, and r is not far from unity, the amplitudes of the tides introduced by friction will be in proportion to the corresponding terms in the force that produces them, and the phase lags will be equal. Thus in 14·21 (11) we have

$$\frac{H}{H_1}=\frac{P}{Q}=\frac{2}{3}\frac{pa}{qb} \qquad \text{.........(15).}$$

The factor q/p is nearly unity, and may be omitted. The ratio b/a arises from the theory before friction was taken into account; hence in calculating it the assumption that the laws involved are linear in the displacements is correct, and therefore

$$\frac{a}{b}=\frac{m/c^3}{m_1/c_1^{\,3}} \qquad \text{.........(16).}$$

Thus (15) gives
$$\frac{H}{H_1}=\frac{2}{3}\frac{m/c^3}{m_1/c_1^{\,3}} \qquad \text{.........(17),}$$

and finally substituting in 14·21 (11) we have

$$\frac{N}{N_1}=\frac{2}{3}\left(\frac{m/c^3}{m_1/c_1^{\,3}}\right)^2 \qquad \text{.........(18)}$$
$$=3\cdot 4.$$

Thus the solar tidal friction is a larger fraction of the lunar in this case than when the friction follows a linear law.

14·3. *The Secular Accelerations of the Sun and Moon.* Let us now consider the effect of the variations in the angular motions on observations of astronomical phenomena. Suppose that n, n_1 and Ω are known from observations over a short interval near time zero, and that from the values found we infer that a known fixed star would cross the meridian of Greenwich at time T if all these quantities were constant. Then the effect of the variation in the speed of the rotation of the earth is to put the earth ahead in time T by an angle $\frac{1}{2}T^2\frac{d\Omega}{dt}$, if we neglect variations in $\frac{d\Omega}{dt}$. The earth's angular velocity being Ω, the meridian of Greenwich will therefore reach a given star in time $T-\frac{1}{2}\frac{T^2}{\Omega}\frac{d\Omega}{dt}$ instead of in time T; in other words, the time of transit of a given star across the meridian is hastened by $\frac{T^2}{2\Omega}\frac{d\Omega}{dt}$. The moon moves among the stars with angular velocity n, and therefore the alteration of the time of the observation makes the moon an angle $\frac{nT^2}{2\Omega}\frac{d\Omega}{dt}$ behind its calculated position when the star transits. On the other hand, the change in its own angular velocity puts it ahead by an angle $\frac{1}{2}T^2\frac{dn}{dt}$ in time T. Altogether, then, it is $\frac{1}{2}T^2\left(\frac{dn}{dt}-\frac{n}{\Omega}\frac{d\Omega}{dt}\right)$

in front of its calculated position when the star transits. Thus, in comparison with the positions inferred for time T, the moon is subject to an advance through an angular distance $\frac{1}{2}T^2\left(\frac{dn}{dt}-\frac{n}{\Omega}\frac{d\Omega}{dt}\right)$. This is equivalent to saying that the moon appears to have, relative to the stars, a steady secular acceleration $\frac{dn}{dt}-\frac{n}{\Omega}\frac{d\Omega}{dt}$. Let us denote this by ν. The sun similarly will have a secular acceleration, which we shall denote by ν_1. Then

$$\nu=\frac{dn}{dt}-\frac{n}{\Omega}\frac{d\Omega}{dt} \qquad \text{.........(1)},$$

$$\nu_1=\frac{dn_1}{dt}-\frac{n_1}{\Omega}\frac{d\Omega}{dt} \qquad \text{.........(2)}.$$

But from 14·2 (2) $$\frac{dn}{dt}=-3n_0\xi^{-4}\frac{d\xi}{dt};\quad \frac{dn_1}{dt}=-3n_{10}\xi_1^{-4}\frac{d\xi_1}{dt} \qquad \text{.........(3)}.$$

Let us now substitute in (1) and (2) the values of the differential coefficients found in 14·2 (3), (4) and (5). Then

$$\nu=-3\frac{M+m}{Mm}\frac{N\xi^{-4}}{c_0^2}+\frac{N+N_1}{C\Omega}n_0\xi^{-3} \qquad \text{.........(4)},$$

$$\nu_1=-3\frac{m_1+M}{m_1M}\frac{N_1\xi_1^{-4}}{c_{10}^2}+\frac{N+N_1}{C\Omega}n_{10}\xi_1^{-3} \qquad \text{.........(5)}.$$

So long as intervals of only a few thousand years are considered, ξ and ξ_1 may be taken equal to unity without sensible error. Let us put

$$\frac{Mm}{M+m}\frac{c_0^2n_0}{C\Omega_0}=\kappa \qquad \text{.........(6)},$$

so that κ is the present value of the ratio of the orbital angular momentum to the rotational angular momentum of the earth. Taking

$$C=0{\cdot}334MA^2 \qquad \text{.........(7)},$$

we have $$\kappa=4{\cdot}82 \qquad \text{.........(8)}.$$

Then (4) reduces to $$\nu=\frac{M+m}{Mmc^2}\{(\kappa-3)N+\kappa N_1\} \qquad \text{.........(9)},$$

where the zero suffixes have now been dropped, since the quantities are nearly equal to their present values.

The ratio of the first term of (5) to the second is

$$\frac{3N_1}{N+N_1}\frac{c^2n}{c_1^2n_1}\frac{m}{M+m}\frac{1}{\kappa}.$$

The first factor is of order unity, the second 10^{-4}, the third 10^{-2}, the last $\frac{1}{5}$. This ratio is therefore very small, and the first term in (5) may be neglected. Then (5) becomes

$$\nu_1=\kappa\frac{M+m}{Mm}\frac{n_1}{n}\frac{N+N_1}{c^2} \qquad \text{.........(10)}.$$

14·31. Now ν and ν_1 are capable of being found from a comparison of present-day astronomical observations with ancient ones. In particular,

it is evident that a determination of the time of the occultation of a star by the moon, or of the conjunction of the moon with a star, will give ν directly. Again, the moon in time T will gain on the sun in longitude by $\frac{1}{2}(\nu - \nu_1) T^2$ in comparison with the calculated motion, and therefore all lunar eclipses will occur earlier by an interval $\frac{1}{2}\frac{\nu - \nu_1}{n - n_1} T^2$. Thus observations of the times of eclipses make it possible to find $\nu - \nu_1$.

The magnitude of a lunar eclipse is determined by the distance of the moon from the node at the time of conjunction or opposition to the sun. The motion of the node has no part arising from tidal friction, and is therefore known from the purely gravitational theory of the moon's motion. The longitude of the moon at the time of eclipse is increased by an amount $-\frac{1}{2}\frac{n}{n - n_1}(\nu - \nu_1) T^2$ on account of the earlier time of the eclipse, but also by $\frac{1}{2}\nu T^2$ on account of its own secular acceleration. Thus the effect of the two accelerations jointly is to increase the moon's longitude at the time of eclipse by $\frac{1}{2}\frac{n\nu_1 - n_1\nu}{n - n_1} T^2$. The magnitudes of the eclipses therefore give $n\nu_1 - n_1\nu$. The times and magnitudes together give ν and ν_1 separately.

The same arguments would apply to a solar eclipse as seen from the centre of the earth; but owing to the secular change in the rate of the earth's rotation the spot on the surface where the magnitude of the eclipse is greatest will be altered. Thus the discussion of solar eclipses is more complicated than that of lunar ones.

It may be noticed, incidentally, that

$$n\nu_1 - n_1\nu = n\frac{dn_1}{dt} - n_1\frac{dn}{dt},$$

and therefore the magnitudes of the eclipses do not involve the change in the rotation of the earth. This result is evident *ab initio*, since they involve only the positions of the sun and moon and the centre of the earth. Again, dn_1/dt has been seen to be very small, so that practically the magnitudes of the eclipses give a direct determination of dn/dt.

Observations of the time of passage of the sun across the equator, when the precession of the equinoxes is known, give the secular acceleration of the sun directly.

Many ancient observations by Greek, Babylonian, Chinese, and Egyptian astronomers have been discussed with a view to the determination of the secular accelerations. The most recent and the most thorough of these discussions is that of Dr J. K. Fotheringham. The following table is taken from his summary* of the information obtainable directly from observations of several different types. His L is equivalent to $\frac{1}{2}\nu$ of this chapter. The unit is a second of arc per century per century.

* *M.N.R.A.S.* **80**, 1920, 578–81.

	Lunar eclipses: Times	Lunar eclipses: Magnitudes	Solar eclipses	Occultations and Conjunctions	Equinoxes
ν	—	—	21·0	21·60 ± 1·40	—
ν_1	—	3·56 ± 0·90	2·0	—	3·66 ± 0·54
$\nu-\nu_1$	17·8 ± 2·6	—	19·0	—	—

Fotheringham has discussed the solar eclipses afresh*, and it is shown by the diagram on p. 123 of his paper that a secular acceleration of the sun of $2''\cdot1/(\text{century})^2$, and one of the moon of $20''\cdot7/(\text{century})^2$ would satisfy them all, but no smaller value of the solar secular acceleration is permissible. It might, however, be as large as $3''\cdot3/(\text{century})^2$, combined with one of the moon of $22''\cdot0/(\text{century})^2$. He decides that on the whole, judging purely on the observational evidence, the most probable values of the secular accelerations are $21''\cdot6/(\text{century})^2$ and $3''\cdot0/(\text{century})^2$; but they satisfy the observations more accurately than the probable errors would lead one to expect, and consequently some latitude of interpretation is permissible.

The lunar theory, however, indicates a secular acceleration of the moon of $12''\cdot2/(\text{century})^2$. The portion attributable to tidal friction is therefore only the unexplained excess, namely about $9''\cdot0/(\text{century})^2$.

14·32. With these values of ν and ν_1 it should be possible to solve 14·3 (9) and (10) for N and N_1 and to find their ratio, and also to substitute in 14·2 (6) and find the rate of dissipation of energy directly, without any hypothesis as to the nature of the tidal friction. It appears, however, that the solar acceleration is uncertain by a larger fraction of its amount than the lunar, and it may be preferable to determine theoretically the ratio of the two accelerations, and thus to infer the solar acceleration from the lunar. Even this cannot, however, be done with much accuracy if the friction varies as the square of the velocity, for squares of the height of the solar tide had to be neglected in 14·23 in inferring the ratio of the frictional couples, and an error of the order of 10 per cent. in the result is therefore to be anticipated.

Now 14·3 (9) and (10) give

$$\frac{\nu}{\nu_1} = \frac{\frac{\kappa-3}{\kappa}N + N_1}{N + N_1}\frac{n}{n_1} \qquad \text{.........(1).}$$

When N/N_1 tends to zero, this tends to $n/n_1 = 13\cdot3$. This is obvious without analysis, for it corresponds to the case where all friction is in the solar tides; and then dn/dt is zero, and dn_1/dt insignificant. Thus the whole of the secular accelerations arise from errors in the time introduced by variations in the rate of rotation, and are therefore in the ratio of the mean motions. Then to a secular acceleration of the moon of $9''\cdot0/(\text{century})^2$ corresponds one of the sun of $0''\cdot7/(\text{century})^2$.

* *M.N.R.A.S.* 81, 1920, 104–26.

When N/N_1 tends to infinity, so that all the friction is in the lunar tides, ν/ν_1 tends to $\frac{\kappa-3}{\kappa}\frac{n}{n_1} = 5{\cdot}0$. This gives

$$\nu_1 = 1''{\cdot}8/(\text{century})^2,$$

which must be the maximum value possible of the solar secular acceleration.

With N/N_1 equal to 5·1, we get

$$\nu/\nu_1 = 6{\cdot}3, \quad \nu_1 = 1''{\cdot}44/(\text{century})^2.$$

With N/N_1 equal to 3·4, we have

$$\nu/\nu_1 = 7{\cdot}2, \quad \nu_1 = 1''{\cdot}26/(\text{century})^2.$$

We notice that the largest possible value of the solar secular acceleration is somewhat less than the lowest permitted by Fotheringham's investigations. This suggests that either an unknown cause is producing a secular acceleration of the sun, or that part of the observed value is error.

We have from 14·3 (9)

$$N\frac{M+m}{Mmc^2}\left\{(\kappa-3)+\kappa\frac{N_1}{N}\right\} = \nu \qquad \text{.........(2)},$$

and from 14·2 (6)
$$-\frac{dE}{dt} = N(\Omega-n)\left\{1+\frac{N_1}{N}\frac{\Omega-n_1}{\Omega-n}\right\} \qquad \text{.........(3)}.$$

Taking
$$\nu = 9''{\cdot}0/(\text{century})^2 = 4{\cdot}5 \times 10^{-24}/\text{sec.}^2,$$
$$N/N_1 = 5{\cdot}1,$$

we have
$$\frac{N}{C} = \frac{73\nu}{1+\frac{8}{3}\frac{N_1}{N}} = 2{\cdot}2 \times 10^{-22}/\text{sec.}^2 \qquad \text{.........(4)},$$

$$-\frac{dE}{dt} = 1{\cdot}5 \times 10^{19} \text{ ergs per second} \qquad \text{.........(5)}.$$

With the same value of ν, but
$$N/N_1 = 3{\cdot}4,$$

we find
$$\frac{N}{C} = 1{\cdot}9 \times 10^{-22}/\text{sec.}^2 \qquad \text{.........(6)},$$

$$-\frac{dE}{dt} = 1{\cdot}39 \times 10^{19} \text{ ergs per second} \qquad \text{.........(7)}.$$

14·4. *Tidal Friction in the Sea: Quantitative Estimates.* Let us now consider where the dissipation of energy takes place. In the open ocean the two components of horizontal velocity u and v, except near the bottom, satisfy differential equations of the form

$$\frac{du}{dt} - 2\omega v = g\frac{\partial}{\partial x}(\bar{\zeta}-\zeta) \qquad \text{.........(1)},$$

$$\frac{dv}{dt} + 2\omega u = g\frac{\partial}{\partial y}(\bar{\zeta}-\zeta) \qquad \text{.........(2)},$$

where x and y are elements of arc in two perpendicular horizontal directions,

ζ is the elevation of the sea-level above its mean position and $\bar{\zeta}$ the height of the equilibrium tide. Also

$$\omega = \Omega \cos \theta \qquad \ldots\ldots\ldots(3).$$

We shall expect that in the absence of close resonance ζ and $\bar{\zeta}$ will be of the same order. Thus u and v will be of order $g\bar{\zeta}/\omega A$, where A is the radius of the earth. Now $g\bar{\zeta}$ is equal to the tidal disturbing potential, and therefore, by 14·21 (6), its maximum value is $\frac{3}{4}\frac{fmA^2}{c^3}$, making ζ of order 26 cm. Thus u and v are of order 1 cm./sec. The frictional force is $k\rho\,(u^2+v^2)$, directed against the resultant velocity, and therefore the rate of dissipation of energy is $k\rho\,(u^2+v^2)^{\frac{3}{2}}$ per unit area, where $k = 0{\cdot}002$ and ρ is the density of water. Thus the dissipation per square centimetre is of order 0·004 erg/sec. The area of the whole ocean being about $3{\cdot}67 \times 10^{18}$ cm.^{2}, the dissipation as a whole must be of the order of 10^{16} ergs/sec., which is a small fraction of that seen to be required to account for the secular accelerations. Thus the chief part of the tidal dissipation cannot arise from friction in the open ocean.

The validity of the above estimate depends on whether the Osborne Reynolds criterion for turbulent motion is satisfied. If the motion were purely viscous, the viscosity being ν, and if the influence of the boundary were considerable up to a distance h from it, the criterion for purely viscous motion is that uh/ν shall be less than 1000. For semidiurnal motions in a viscous medium,

$$h^2 = \nu/\omega,$$

so that

$$uh/\nu = u/(\nu\omega)^{\frac{1}{2}} = 800.$$

Thus the criterion is on the verge of being satisfied, and it is known that the dissipation by pure viscosity cannot in any case exceed that by turbulence when the velocity is just great enough to satisfy the Reynolds criterion. The order of magnitude of the above estimate is therefore not too low in any case.

14·41. It was, however, assumed explicitly in the last section that the conditions are such that ζ is not great compared with $\bar{\zeta}$. If the form of the bottom is such that the tides are highly magnified by resonance or by the shallow water effect, the argument will not hold. Such circumstances cannot exist in the ocean as a whole, but they may well hold locally. Indeed places where tidal currents have velocities very much greater than the 1 cm./sec. inferred for the open ocean in the last section are known to everybody. In such places the dissipation per unit area must much exceed that in the open ocean, especially since the rate of dissipation per unit area is proportional to the cube of the velocity. The question is, whether the increase in the rate of dissipation per unit area is enough to compensate for the limited area of the regions concerned and make the total dissipation in them exceed that in the open ocean.

The question was answered in the affirmative by G. I. Taylor*, who determined the rate of dissipation of energy in the tides in the Irish Sea in two independent ways. One method is the natural one of estimating the velocities all over the area considered, and hence inferring the dissipation per unit area. On integrating this over the whole region we find the total dissipation.

The alternative method is to find the rate of performance of work on the sea by the ocean and the moon together. This is found to be, on the average, positive. The energy in the sea is, however, not increasing steadily, and therefore the energy supplied from outside must be dissipated as fast as it is received. Thus the rate of dissipation can be found. If we consider a straight line across the sea as the boundary, and take the axis of y along this line, the velocity across it is u. Let D be the depth of the water at any point. The pressure at any depth z below the mean position of the surface is $g\rho\,(z+\zeta)$, since the depth below the actual free surface is $z+\zeta$. Thus the entering water does work at a rate $g\rho\,(z+\zeta)\,u$ per unit area of the section. Integrating this with regard to z from $z=-\zeta$ to $z=D$, supposing u independent of z, we see that the rate of performance of work by the water entering vertically below an element of the boundary of unit length is

$$\int_{-\zeta}^{D} g\rho\,(z+\zeta)\,u\,dz = \tfrac{1}{2}g\rho\,(D+\zeta)^2\,u \qquad \text{.........(1).}$$

In addition the entering water brings in its own energy. Taking the mean sea-level as the zero of potential, we see that the mean potential within a column of unit cross section extending to height ζ and depth D is $-\tfrac{1}{2}g\,(D-\zeta)$. The mass of such a column is $\rho\,(D+\zeta)$. The potential energy is therefore $-\tfrac{1}{2}g\rho\,(D^2-\zeta^2)$. Now the horizontal cross section of the column of water that crosses in unit time unit length of the boundary is u. Thus the potential energy of the entering water enters at a rate

$$-\tfrac{1}{2}g\rho\,(D^2-\zeta^2)\,u \qquad \text{.........(2)}$$

per unit length of the boundary.

The kinetic energy of the entering water evidently contributes energy to the sea at a rate

$$\tfrac{1}{2}\rho\,(D+\zeta)\,(u^2+v^2)\,u \qquad \text{.........(3)}$$

per unit length.

Combining these three sources of energy, we find for the whole rate of inflow of energy

$$g\rho u\,(D\zeta+\zeta^2)+\tfrac{1}{2}\rho\,(D+\zeta)\,u\,(u^2+v^2) \qquad \text{.........(4)}$$

per unit length of the boundary.

Now in general ζ is small compared with D. Also u and v are in general of the order of $c\zeta/D$, where c is the velocity of a tidal wave in water of depth D. But

$$c^2 = gD,$$

so that $D\,(u^2+v^2)$ is of order $g\zeta^2$. Thus much the largest part of the

* *Phil. Trans.* 220 A, 1919, 1–33.

entering energy is contributed by the term $g\rho uD\zeta$. On integration along the boundary the whole rate of transfer of energy across the boundary is seen to be

$$\int g\rho Du\zeta\, dy \qquad \text{.........(5)}$$

taken along the boundary from end to end. If this again is integrated with regard to the time through a whole period of the motion, we shall obtain the total inflow of energy across the boundary during a period; and if these results are added for all the boundaries of the sea, the total rate of inflow of energy into the sea may be found.

We require in addition the work done by the moon. The disturbing potential due to the moon being U, the work done by the moon when an element of the ocean surface of area dS is raised through a height $d\zeta$ is $U\rho\, d\zeta\, dS$. Putting

$$U = g\bar{\zeta} \qquad \text{.........(6),}$$

so that $\bar{\zeta}$ is the height of the equilibrium tide, we see that the work done by the moon on the sea in a period is

$$\iint g\rho\bar{\zeta}\frac{d\zeta}{dt}\, dt\, dS \qquad \text{.........(7),}$$

where the integral with regard to t is to be taken over a period, and that with regard to S over the whole area of the sea. This added to the result of the integration of (5) through a period gives the whole supply of energy to the sea; and this must be equal to the dissipation of energy in the sea during a period.

14·411. Taylor found the rate of dissipation of energy in the Irish Sea by these two methods. He found that energy enters through St George's Channel and the North Channel at a mean rate of $6{\cdot}4 \times 10^{17}$ ergs per second. The rate of performance of work on the sea by the moon is, on an average, about $-4{\cdot}3 \times 10^{16}$ ergs per second. The remainder, about 6×10^{17} ergs per second, is absorbed by tidal dissipation. The alternative method, in which the dissipation all over the area is estimated simply from the velocity distribution, gave $5{\cdot}2 \times 10^{17}$ ergs per second, agreeing with the other within the limits of error of the determinations of velocity used.

The dissipation in the Irish Sea alone is therefore about sixty times what was found for the open ocean as a whole. The conjecture that the greater velocity of the currents may more than make up for the smaller area therefore proves to be correct. The rate of dissipation is indeed 1/20 of that required to account for the whole of the secular acceleration of the moon, and it was therefore suggested by Taylor that the dissipation of energy in the whole of the shallow seas of the earth may be enough to explain the whole of the secular acceleration.

14·412. It may be noticed, however, that the above estimate for the Irish Sea refers definitely to spring tides, when the tidal currents are at a maximum. To find the mean dissipation throughout the lunar month

a correcting factor must therefore be applied. Let us then proceed to estimate this factor. Let θ be the phase of the lunar tide, and $r\theta$ that of the solar tide. Put $r = 1 - s$, so that s is $\frac{1}{29}$. Let A be the amplitude of the lunar current and $A\nu$ that of the solar current. Then the total current is

$$A\{\cos\theta + \nu\cos(1-s)\theta\} = A(1 + 2\nu\cos\theta + \nu^2)^{\frac{1}{2}}\cos\left(\theta - \tan^{-1}\frac{\nu\sin s\theta}{1+\nu\cos s\theta}\right),$$

which is now expressed as a simple harmonic motion with a slowly varying amplitude and period. The amplitude at springs is $A(1+\nu)$. The rate of dissipation is proportional to the cube of the current, and therefore that over a lunar day is proportional to the cube of the amplitude during that day. Thus the ratio of the mean dissipation to the dissipation at springs is the average of

$$(1 + 2\nu\cos s\theta + \nu^2)^{\frac{3}{2}}/(1+\nu)^3.$$

If ν^6 be neglected, the numerator of this is $1 + \frac{9}{4}\nu^2 + \frac{9}{64}\nu^4$. Assuming that the velocities vary in proportion to the vertical ranges, we have

$$\frac{1+\nu}{1-\nu} = 2{\cdot}3,$$

$$\nu = 0{\cdot}39,$$

$$\frac{1 + \frac{9}{4}\nu^2 + \frac{9}{64}\nu^4}{(1+\nu)^3} = 0{\cdot}51.$$

Thus the correcting factor is about 0·5, and the number of Irish Seas required to account for the secular acceleration of the moon is between 40 and 50.

14·413. Taylor's methods have been extended by the present writer* to include most of the shallow seas of the globe. The results are as follows. All refer to spring tides. The unit is 10^{18} ergs per second.

European waters:		Asiatic waters:		North American waters:	
Irish Sea	0·6	South China Sea	Small	Hudson Strait	0·2
English Channel	1·1	Yellow Sea	1·1	Hudson Bay	Small
North Sea	0·7	Sea of Okhotsk	0·4	Fox Strait	1·4
		Bering Sea	15·0	Bay of Fundy	0·4
		Malacca Strait	1·1		

The other seas of the globe appear unlikely to contribute much to the dissipation. In Europe, the Mediterranean, Baltic, and White Sea are so narrow at their entrances that little tide can enter, and consequently the tidal currents within them are unable to contribute much to the dissipation. The Bay of Biscay is too deep to magnify the currents much. In Asia, the South China Sea contributes little because the tide in it is almost wholly diurnal; the disturbing potential that produces the diurnal tide is small, arising from the inclination of the moon's orbit to the equator, and dissipation in it contributes nothing to the secular acceleration.

* *Phil. Trans.* 221 A, 1920, 239–64.

Australian waters contribute little, for similar reasons; while the Gulf of Mexico resembles the Mediterranean.

In the cases of the Yellow Sea and Malacca Strait it has been possible to apply both of Taylor's methods of calculating the dissipation, and concordant results have been obtained. In the Bay of Fundy the currents vary so much from point to point that only the method based on the inflow of energy is useful. In the others it has been necessary to rely wholly on the estimates based on the velocity. The results, especially those for the Bering Sea, the most important region of all, are therefore subject to revision as more accurate knowledge of the tides and tidal streams becomes available.

The total dissipation is found to be 22×10^{18} ergs per second at spring tides. Applying the correcting factor 0·51 to obtain the average dissipation, we find $1{\cdot}1 \times 10^{19}$ ergs per second, 80 per cent. of what is required to account for the whole of the secular acceleration of the moon. It would by itself give a secular acceleration of 7″ per century per century.

14·414. The agreement between the dissipation in shallow seas and that necessary to account for the lunar secular acceleration is much closer than the data would entitle us to expect. Two-thirds of that found takes place in the Bering Sea; this determination may be in error by half its amount if the estimates of the currents there are systematically wrong by 25 per cent., a possible contingency. What we are entitled to assert, however, on the evidence before us, is that the dissipation is certainly enough to account for a large fraction of the secular acceleration, and that there is nothing to prove that it is incapable of accounting for the whole of it*. In particular, any portion of the secular acceleration can be attributed to bodily tidal friction only as a result of independent proof that bodily friction must produce such an amount; there is no insufficiency in the theory of tidal friction in shallow seas that can justify such a course.

It is uncertain whether the dissipation in any other coastal regions is notable in comparison with those already considered. The only partially enclosed seas not treated so far are some of those in the North-West Passage. There is an extensive shallow region off the coast of Patagonia, but it is in no way enclosed, being perfectly open to the Atlantic. Thus it is difficult to make any reliable inference about the currents in it. Many records of tidal currents along the coast are given in the Admiralty Pilots, but all of them seem to refer to currents up rivers or near their mouths, or to currents over bars and shoals; in either case the general currents must be magnified. There seem to be no data about currents more than a few miles out to sea.

The dissipation over local shallows like shoals and bars, and in narrow bays and straits, has been systematically ignored in this discussion, except

* A discussion based on substantially the same data, and giving comparable results, is given by W. Heiskanen in *Ann. Acad. Scient. Fennicae*, 18 A, 1921, 1–84.

where it has been automatically taken into account in the determination of the excess of the entering over the issuing energy. The chief reason is the utter impossibility of finding it from our present data. It can be proved that variations of depth across a channel do not affect the longitudinal currents much, so that shallows long compared with their width will not affect the result; but roughly circular shoals, and shallow regions extending right across a channel, may contribute an appreciable amount. The agreement found in the cases of the Irish Sea, the Yellow Sea, and the Strait of Malacca between the results of the two methods of computing the dissipation shows that at any rate the shoals do not contribute an amount comparable with that in the regions as a whole; for one method systematically includes the effect of shoals, and the other systematically omits it.

The fjords of the west coasts of Norway, Greenland, and North and South America, and the Scottish lochs, are innumerable. They are, however, mostly very deep, and strong tidal currents are therefore not to be expected over much of their area. Strong currents do, however, occur locally in them. No attempt has yet been made to compute the dissipation in these regions, but it is probably small.

Along the open shore, again, there must be some dissipation; the strong currents usually do not extend more than a few miles out to sea, but they exist along a very long stretch of coast, and the aggregate dissipation in them may be appreciable.

14·42. *Bodily Tidal Friction.* Let us proceed now to consider the possible effects of bodily imperfection of elasticity. We saw in Chapter IX that imperfection of elasticity in a solid may be manifested in two ways, either by a slowness of elastic recovery after strain, called elastic afterworking, or by steady deformation under stress, called plasticity. It was seen to be possible for plasticity to appear only when the stress exceeded a certain value, which was called the 'strength'; but it was realised that the strength might be zero. It is possible for elastic afterworking and plasticity to be present in the same substance.

No attempt was made to state quantitative laws of these properties, and it appears probable that in actual substances the laws are very complex. In the case of bodily tides in the earth, however, we can make some progress. The amplitude of the bodily tide is, like that of the ocean tide, of the order of a metre. Thus each part of the earth is stretched during the lunar day by amounts comparable with 10^{-7} of its linear dimensions. It is incredible that if any terms in the friction depend on second and higher powers of the displacements, they can be appreciable when the displacements are as small as this. On the other hand, if imperfection of elasticity arises only when the stresses exceed a certain amount, and the frictional stresses thereafter remain constant, as in solid friction and some

forms of plasticity, we cannot expect the strength to be so small as this would indicate, and therefore the substance would be perfectly elastic for the small stresses involved in bodily tides.

14·421. If then any imperfection of elasticity is present, we must expect it to depend on the first power of the displacements and their rates of change. It is easy to obtain two plausible forms of imperfection of elasticity that show the characteristic features of plasticity and elastic afterworking respectively. We have seen that if P represents a tangential stress, and E the corresponding distortional strain, the relation between them in a perfectly elastic solid is

$$P = \mu E \qquad \text{.........(1)},$$

where μ is the coefficient of rigidity. Now it may happen that the distortion is changing, and if so, the stress is not only maintaining the distortion at its temporary value, but also is producing new distortion. There may be a further resistance to the latter process; it vanishes with the rate of distortion, by definition, and, since it must be linear in the distortion and its rates of change, must therefore be proportional to the rate of increase of distortion. Thus we must have

$$P = \mu \left(E + \mathrm{t}_2 \frac{dE}{dt} \right) \qquad \text{.........(2)},$$

where t_2 is a constant with the dimensions of a time. We see that if the stress is removed the strains diminish exponentially to zero, falling to e^{-1} of their original values in time t_2. If the system is in a state of steady strain, we shall have $$P = \mu E$$ as before; thus for motions periodic in times long compared with t_2 the substance will behave as if perfectly elastic. For motions with periods short compared with t_2, on the other hand, the term in $\mathrm{t}_2 \dfrac{dE}{dt}$ will much exceed that in E. If then μ is finite, any finite value of E would make the right side of (2) greater than the left. E must therefore be zero. Thus when the motions have short periods, and μ is finite, the substance behaves as if perfectly rigid. Thus this type of imperfection of elasticity implies not weakness, but additional stiffness. For this reason it will be referred to as 'firmoviscosity.' It is evidently a particular case of elastic afterworking.

On the other hand, if $\mu \mathrm{t}_2$ is finite but μ zero, the relation takes the form

$$P = \nu \frac{dE}{dt},$$

which is the characteristic form of the stress-strain relation in a viscous fluid. If finally $\mu \mathrm{t}_2$ is also zero, P is necessarily zero, and we have a perfect fluid.

14·422. On the other hand, it is possible that the imperfection of elasticity may be of plastic type; that is, if the stress is kept constant, the

strain will increase steadily. Then E will contain two parts. The first, P/μ, arises from simple elastic strain. The second must be supposed to increase at a rate proportional to P. Thus we must have

$$\mu E = P + \frac{1}{t_1}\int P\,dt \qquad \text{.........(1).}$$

In this case a sudden stress gives an immediate strain P/μ as in a perfectly elastic solid, but the strain then proceeds to increase at a uniform rate $P/\mu t_1$ so long as the stress is kept constant. When the stress is removed, the solid does not return to its original configuration, however long it may be left; it retains a permanent set measured by $PT/\mu t_1$, where T is the time of application of the stress.

This type of imperfection of elasticity is called 'elasticoviscosity.' The behaviour of elasticoviscous solids under periodic stress is fundamentally different from that of firmoviscous ones. If the period is long compared with t_1, the second term in (1) will much exceed the first, and the stress-strain relation may be rewritten

$$\mu t_1 \frac{dE}{dt} = P,$$

which is the form appropriate to a viscous liquid. Thus the elasticoviscous solid approximates to a viscous liquid for any long period forces, whatever its rigidity may be. The firmoviscous solid never approximates to a liquid when the rigidity is finite, and behaves as if perfectly elastic for long-period stresses.

If, on the other hand, the period is short compared with t_1, the second term in (1) is small, and the elasticoviscous solid behaves as if perfectly elastic. In the corresponding case a firmoviscous solid behaves as if perfectly rigid.

14·423. It is possible to combine both types of imperfection of elasticity in one substance. For, if we have

$$\mu\left(E + t_2\frac{dE}{dt}\right) = P + \frac{1}{t_1}\int P\,dt \qquad \text{.........(1),}$$

the substance will follow the firmoviscous law if t_1 is infinite, and the elasticoviscous one if t_2 is zero. Such a substance will flow indefinitely with long-continued stresses, but the partial recovery on release will be gradual, whereas in a purely elasticoviscous substance it is instantaneous. Under quickly changing stresses the material will behave as if perfectly rigid, like a firmoviscous substance.

We see that if any problem of elastic strain has been solved for a perfectly elastic solid, the behaviour of an imperfectly elastic one, so long as squares of the displacements can be neglected, can be inferred by simply writing

$$\mu\left(1 + t_2\frac{d}{dt}\right)\Big/\left(1 + \frac{1}{t_1\,(d/dt)}\right) \text{ for } \mu \qquad \text{.........(2).}$$

In particular, the behaviour of a firmoviscous solid can be found by replacing μ by $\mu\left(1 + t_2\frac{d}{dt}\right)$, and that of an elasticoviscous one by replacing μ by $\mu\Big/\left(1 + \frac{1}{t_1\,(d/dt)}\right)$.

The elasticoviscous law was formulated by J. Clerk Maxwell*, while the firmoviscous one† was suggested to me privately by Sir J. Larmor.

14·424. Let us now proceed to an estimate of the values of t_1 and t_2 necessary to account for the secular acceleration of the moon on the hypotheses of elasticoviscosity and firmoviscosity respectively. Suppose that in a system with one degree of freedom a periodic force $E \cos pt$ produces a displacement $A \cos(pt - \alpha)$. Then the rate of dissipation of energy is $pEA \cos pt \sin(pt - \alpha)$, and the total dissipation over a period is $\pi EA \sin\alpha$. On the other hand the potential energy due to the external force at the time of greatest displacement is $EA \cos\alpha$. Thus the fraction $\pi \tan\alpha$ of the maximum energy is dissipated in every period.

In the case of the bodily tides in the earth, the external potential is the disturbing potential due to the moon, and the maximum work done by it on unit area of the earth's surface is $g\rho\bar{\zeta}\zeta$, where ρ is the mean density, $\bar{\zeta}$ is the height of the equilibrium tide and ζ the vertical displacement at the surface. Taking ζ and $\bar{\zeta}$ on the equator equal to 25 cm and ρ equal to 3, as we need only consider orders of magnitude, we find that the maximum potential energy per sq. cm. at any point of the surface is $1{\cdot}9 \times 10^6 \sin^4\theta$ ergs, whose mean value over the surface is $1{\cdot}0 \times 10^6$ ergs. The rate of dissipation of energy in the earth, if it were enough to account for the whole of the secular acceleration, would be $1{\cdot}5 \times 10^{19}$ ergs per second. Hence the dissipation in 12 hours in a cone whose vertex is at the centre of the earth and whose base is a square centimetre of the surface is $1{\cdot}3 \times 10^5$ ergs. We have therefore

$$\pi \tan\alpha = 0{\cdot}13 \qquad \text{.........(1)},$$

giving

$$\alpha = 0{\cdot}04 \qquad \text{.........(2)}.$$

Thus the bodily tides should lag by about 3° in phase; but since the phase of the disturbing potential is twice the difference in longitude between the moon and the point considered, it appears from 14·1 that the point of maximum elevation is $1\frac{1}{2}$ degrees of longitude, or six minutes of time, in front of the moon.

Now consider a system acted upon by a restoring force which would make the free period $2\pi/n$. Then the motion under a disturbing acceleration $E \cos pt$ is given by

$$\ddot{x} + n^2x = E\cos pt \qquad \text{.........(3)},$$

and if the free period is short compared with the period of the disturbing force, as is true in the case of the bodily tide, we can omit the first term and write simply

$$n^2x = E\cos pt \qquad \text{.........(4)}.$$

* *Phil. Mag.* 35, 1868, 134. † *M.N.R.A.S.* 77, 1917, 449–56.

Now in this case the tendency of the earth to readjust itself to a symmetrical form is due partly to elasticity and partly to gravity; thus n^2 will be the sum of two positive parts arising from these two properties of matter. In our discussion, however, it will not be necessary to include the effect of gravity, for it can be seen from the work of Kelvin, Darwin, Love and others that it does not affect the order of magnitude of the result. Then n^2 will be proportional to μ, the rigidity.

If now imperfection of elasticity is taken into account, μ must be replaced by the operator given in 14·423 (2). The solution is then to be obtained by replacing $E \cos pt$ on the right by $Ee^{\iota pt}$, d/dt on the left by ιp, solving, and taking for the displacement the real part of the result. Thus we can write

$$n^2 \frac{1 + \mathfrak{t}_2 \iota p}{1 + \dfrac{1}{\mathfrak{t}_1 \iota p}} x = Ee^{\iota pt} \qquad \text{.........(5).}$$

Now x will be of the form $Xe^{\iota(pt-\alpha)}$, where X is real. On taking logarithms of both sides, and equating imaginary parts, we have

$$\tfrac{1}{2}\pi + \tan^{-1} p\mathfrak{t}_2 - \tan^{-1} p\mathfrak{t}_1 = \alpha \qquad \text{.........(6).}$$

If elasticoviscosity is absent, so that $\mathfrak{t}_1$ is infinite, we have

$$p\mathfrak{t}_2 = \tan \alpha,$$

and for a semidiurnal tide, with $\alpha = 0{\cdot}04$, this gives

$$\mathfrak{t}_2 = 270 \text{ secs.}$$

If firmoviscosity is absent, so that $\mathfrak{t}_2$ is zero, we have

$$p\mathfrak{t}_1 = \cot \alpha,$$

and $$\mathfrak{t}_1 = 2 \text{ days.}$$

If both are present, $\mathfrak{t}_2$ will evidently have to be less than 270 secs., and $\mathfrak{t}_1$ will have to be greater than 2 days*.

14·425. It is possible to use these results to obtain further information about the physical properties of the interior of the earth. Consider the motion of a distortional earthquake wave in an imperfectly elastic homogeneous medium. The displacement at a point remote from the focus is proportional to $\dfrac{1}{r} \cos \kappa (c_2 t - r)$, where $2\pi/\kappa$ is the wave length. If p be the period of the wave, we can write this

$$\frac{1}{r} \cos p\left(t - \frac{r}{c_2}\right), \text{ or } R\left(\frac{1}{r} e^{\iota pt} e^{-\iota p r/c_2}\right),$$

where $R(X)$ denotes the real part of X. But we have

$$c_2{}^2 = \mu/\rho,$$

where μ is the rigidity and ρ the density. Hence the motion in an imperfectly rigid body is to be found by writing instead of c_2

$$c_2' = \left(\frac{\mu}{\rho}\right)^{\frac{1}{2}} \left(\frac{1 + \mathfrak{t}_2 \iota p}{1 + \dfrac{1}{\iota p \mathfrak{t}_1}}\right)^{\frac{1}{2}}.$$

* The complete theory, given in *M.N.R.A.S.* **75**, 1915, 648–658, and **77**, 1917, 449–456, makes $\mathfrak{t}_1 > 3$ days, $\mathfrak{t}_2 < 250$ secs.

Now in an earthquake wave the period is of the order of 20 seconds. If then the tidal dissipation is to be attributed to firmoviscosity pt_2 will be of order 100, and pt_1 infinite, and an approximation to c_2' will be $7c_2(1+\iota)$. Again, we have

$$e^{-\iota pr/c_2} = \exp - \frac{\iota pr}{c_2} \frac{1}{7(1+\iota)} = \exp \frac{1}{14} \frac{pr}{c_2} (\iota - 1).$$

Thus the displacement will contain a factor $e^{-\frac{1}{14}pr/c_2}$. If $c_2 = 4$ km./sec., $p = 2\pi/20$ secs., this factor would be $1/e$ if $r = 170$ km. Thus distortional waves could not penetrate to more than a few hundred kilometres from their origin, whereas we know that actually they can penetrate most of the earth and produce observable disturbances on the other side. It follows that the firmoviscosity of the earth in the region where earthquake waves travel must be a small fraction of that required to account for the secular acceleration; only a tiny fraction of the dissipation can be explained by firmoviscosity in this region.

Elasticoviscosity of the amount required to account for the secular acceleration would give no appreciable damping of earthquake waves. On the other hand, we have seen that if a motion has a period long compared with t_1, the substance considered will behave like a perfect fluid. Now we have seen that if any considerable fraction of the tidal dissipation is to be attributed to elasticoviscosity, t_1 cannot much exceed 2 days. Thus in any motion with a period much longer than 2 days the earth would behave like a perfect fluid. Now we shall see in the next chapter that the earth has a motion called the Eulerian nutation, or the 14-monthly variation of latitude, whose existence depends wholly on the earth's not behaving as a fluid. The existence of elasticoviscosity to an important extent therefore appears improbable.

We therefore infer that imperfection of elasticity within the solid earth can probably contribute nothing appreciable to the secular acceleration. The failure of the centre of the earth to transmit earthquake waves may perhaps be attributable to firmoviscosity there, but it is quite certain that firmoviscosity in the outer regions can make no important contribution. The regions of the crust down to the asthenosphere certainly do not show plastic yield to small forces in a few days, but elasticoviscosity near the centre may be appreciable; how important it can be may possibly be estimated from the theory of the variation of latitude.

14·5. *Tidal Friction in the Past.* If we adopt the results of 14·32, we have

$$\frac{d\Omega}{dt} = -\frac{N + N_1}{C}$$

$$= -2{\cdot}5 \times 10^{-22}/\text{sec.}^2$$

The present value of Ω is $7{\cdot}3 \times 10^{-5}$/sec. Thus Ω changes by 10^{-5} of its amount in 3×10^{12} secs., or 10^5 years. The day has therefore probably

lengthened by a second in the last 120,000 years. Thus tidal friction, historically speaking, is a slow process. On the other hand, if we consider the change since the oldest known rocks were formed, and suppose tidal friction to have operated ever since at its present rate, we find that the period of rotation $1{\cdot}6 \times 10^9$ years ago must have been only about 0·84 of our present day. The ellipticity of the earth's surface corresponding to such a rate of rotation is about $\frac{1}{210}$.

Now we saw in Chapter XIII that the earth is probably adjusted to the hydrostatic state so accurately that the ellipticities of the strata of equal density within it do not deviate from their hydrostatic values by more than one part in 150. Yet here we have reason for believing that a change of about one part in four has taken place in geological time; for though the distribution of shallow seas has certainly changed, their effect must always have been in the same direction, and we may provisionally suppose that the hypothesis of uniformity gives the correct order of magnitude. It appears, therefore, that the earth has probably adjusted itself to the hydrostatic form at several epochs during geological time. The earth, originally highly elliptical, became too elliptical for its rotation when this slackened owing to tidal friction, and when the stress-differences in the interior reached the strength the interior adjusted itself to the hydrostatic state, the ellipticity being thereby reduced. The adjustment of the interior might be almost or quite a continuous process, on account of the weakness of the material. The strong outer layers, on the other hand, would not give way until the stresses had accumulated to a certain finite amount, and when they ultimately gave way mountain ranges would probably be formed. The compression available from this source is appreciable, but not so great as that attributable to thermal contraction. The excess of the equatorial radius over the mean is $\frac{1}{3}A\epsilon$, where A is the radius and ϵ the ellipticity. When the ellipticity changes from 1/210 to 1/295, the circumference is accordingly found to decrease by about 14 km. This is not of much geological importance; but if it should appear that tidal dissipation is at present unusually small, the change in the rotation of the earth will have to be regarded as an important factor in mountain formation*.

14·51. If we attempt to take the extrapolation further back still, we must have recourse to equations 14·2 (3), (4), and (5). The couple N has been seen to be proportional to $(m/c^3)\,H \sin 2\epsilon$, and if we suppose ϵ to remain constant and H to be proportional to (m/c^3), as is not unreasonable, N will be proportional to c^{-6} or ξ^{-12}. It therefore increases very rapidly as the moon's distance from the earth diminishes. We shall denote its present value by N_0. The solar tidal friction is at present a small fraction of the lunar, and cannot have changed much, since the sun's distance

* Compare also a forthcoming paper by R. Stoneley in *M.N.R.A.S., Geophys. Suppt.*

has hardly varied. Thus it must have been unimportant in the past in comparison with lunar tidal friction. Hence we can omit equation 14·2 (4) and drop N_1 in 14·2 (5). Then

$$\frac{Mm}{M+m} c_0{}^2 n_0 \frac{d\xi}{dt} = N_0 \xi^{-12} \qquad \text{.........(1)},$$

$$C \frac{d\Omega}{dt} = -N_0 \xi^{-12} \qquad \text{.........(2)},$$

or if we introduce κ as in 14·3 (6), (1) becomes

$$\kappa C \Omega_0 \frac{d\xi}{dt} = N_0 \xi^{-12} \qquad \text{.........(3)},$$

whence

$$\tfrac{1}{13}(1-\xi^{13}) = -\frac{N_0 t}{\kappa C \Omega_0} \qquad \text{.........(4)}.$$

Substituting our adopted values of N_0/C, κ, and Ω, we find for the time when ξ was 0·8, and therefore the distance of the moon 240,000 km.,

$$t = -4 \times 10^9 \text{ years} \qquad \text{.........(5)}.$$

The time taken by the moon to recede from its closest approach to the earth to 240,000 km. from it would not exceed 1/20 of this. Thus the theory of tidal friction suggests an age of the moon of the order of 4×10^9 years. This is about three times the age of the oldest rocks found from radioactivity, but evidently cannot be used to do more than suggest the order of magnitude of the age of the moon.

Equations (2) and (3) together give

$$\frac{d}{dt}(\kappa \Omega_0 \xi + \Omega) = 0 \qquad \text{.........(6)}.$$

This is a form of the equation of angular momentum of the system. Evidently if Ω was once $\frac{4}{3}\Omega_0$, $\kappa\xi$ must have been $4{\cdot}82 + 1 - \dfrac{\Omega}{\Omega_0} = 4{\cdot}49$, making ξ equal to 0·93, the distance of the moon only about 27,000 km. less than at present, and the length of the month 22 of our present days. Thus the distance and period of the moon have, on the hypothesis of uniformity, not changed by large fractions of themselves during geological time.

14·52. *The Future of the Earth-Moon System.* We saw in 14·1 that a necessary condition for the existence of tidal friction is that the earth's surface shall revolve in such a way that the moon's apparent altitude, as seen from each point of the earth's surface, is continually varying. If the earth always kept the same face turned towards the moon, the tides would settle down exactly under the moon and opposite to it, and would produce no tidal frictional couple. Thus N vanishes with $\Omega - n$. In no other circumstances can N vanish. In the primitive state of the system the length of the month was probably slightly greater than that of the day, and therefore tidal friction made the moon recede; the moon will continue to recede until $\Omega - n$ vanishes, when tidal friction will vanish again.

Now when this happens Ω will be equal to n, and therefore to $n_0 \xi^{-3}$. By 14·51 (6), $\kappa\xi + \frac{\Omega}{\Omega_0}$ must still have its present value 5·82. Hence

$$\kappa\xi + \frac{n_0}{\Omega_0}\xi^{-3} = 5{\cdot}82 \qquad \text{.........(1).}$$

The smaller root of this equation is $\xi = 1/5{\cdot}1$, corresponding to the primitive state of the system; the larger is $\xi = 1{\cdot}20$. The former makes the common period of rotation and revolution 4·8 hours; the latter makes it 47 days. Since $1{\cdot}2^{13} - 1 = 10$ nearly, the time needed to approach this latter state will be of order 5×10^{10} years.

The matter will not be closed when this state is reached, however. The solar tides will continue to operate on the earth and lengthen its period of rotation. Thus the period of rotation of the earth will come to be longer than the period of revolution of the moon. A reference to 14·1 will show that when this happens the further course of the evolution is considerably altered. The fixed point P on the earth's surface will move round more slowly than the point where the moon is in the zenith, and therefore the high tides, while still occurring after the moon has passed the zenith or nadir, will be on the opposite sides of AB from C and D. Thus their effect on the moon will be to retard its revolution and make it return to the earth, and the earth's rotation will at the same time be accelerated. Thus the moon will gradually return towards the earth, the earth's rotation meanwhile being retarded by the solar tides and accelerated by the lunar ones. This process will continue till the moon is at last dragged down to within Roche's limit, when it will be broken up by the action of the tides raised in it by the earth; it will then ultimately form a system like Saturn's ring, but much more massive.

14·6. *The History of the Moon's Rotation.* Let us now return to equation 14·21 (10). The height of the equilibrium tide is U/g, and its amplitude is therefore $\frac{3}{4}\frac{fmA^2}{c^3 g}$, or $\frac{3}{4}\frac{mA^4}{Mc^3}$. If then the tide has approximately its equilibrium height, we shall have

$$-\frac{d\Omega}{dt} = \frac{N}{C} = \frac{18}{5}\pi f\rho \sin 2\epsilon \left(\frac{m}{M}\right)^2 \left(\frac{A}{c}\right)^6 \qquad \text{.........(1).}$$

We shall have recourse to this expression again in considering the tides in the planets.

If ω denotes the moon's angular velocity of rotation, it appears from (1) that

$$\frac{d\omega/dt}{d\Omega/dt} = O\left\{\left(\frac{M}{m}\right)^4 \left(\frac{a}{A}\right)^6\right\} \qquad \text{.........(2),}$$

provided elasticity is not so great as to affect the order of magnitude of the height of the tide, that the densities are of the same order, and the phase lags of the tides also. With

$$M/m = 81, \quad a/A = \tfrac{3}{11},$$

we have
$$\frac{d\omega/dt}{d\Omega/dt} = 17{,}000.$$

The hypothesis that the elasticity of the moon may be neglected is valid so long as the moon was largely fluid. The densities are actually of the same order, and the phase lags will be comparable if the departures of the angular velocities of rotation from that of revolution are comparable. Thus if initially the earth and moon rotated at the same rate, this being somewhat different from the mean motion of the system, the rotation of the moon would approach the mean motion 17,000 times as fast as that of the earth would. The moon would thus be brought to present the same face always towards the earth before the rotation of the earth had been appreciably affected by tidal friction.

The moon would be unable to retain any atmosphere or water vapour, on account of its low gravitative power. The absence of the blanketing effect of the atmosphere would enable it to solidify somewhat earlier than the earth. Henceforth any tidal friction in the moon must have arisen from imperfection of elasticity, for there can have been no seas upon it. The thermal history of the moon must have been very similar to that of the earth, and we should therefore expect the departure of its interior from perfect elasticity to be quantitatively comparable with that of the earth. Again, the elastic tide in the moon must be much smaller than the hydrostatic equilibrium tide so far considered, having about $\frac{1}{50}$ of the amplitude. On both grounds the value of $d\omega/d\Omega$ after the solidification of the moon must have become very much smaller; it may well, however, still be as great as 100 for equal lags.

This therefore supplies us with the required explanation of the fact that the moon always keeps the same face towards the earth. The tides raised in the moon by the earth produced such friction that they made the moon's periods of rotation and revolution equal at a very early stage in its history. It is possible that ever since then, if the recession of the moon from the earth, or any internal change in the moon, made either of these periods vary, bodily tidal friction in the solid moon would commence afresh and restore the equality.

14·61. We have, however, seen that there is no reason to suppose bodily tidal friction in the earth to be perceptible, and accordingly there is no great prior probability for a hypothesis that requires it to be continually available in the moon to adjust the moon's rotation to any changes that may have occurred. Let us consider, then, what would happen if the moon was actually free from internal friction. Take an axis OA in the ecliptic, fixed in direction. Let the longest axis of the moon, which points nearly to the earth, make an angle ϕ with OA, and let the line joining the centres of the earth and moon make an angle θ with OA. Neglect the inclination of the moon's orbit and equator to the ecliptic. Put

$$\phi = \theta + \psi \qquad \text{.........(1)},$$

so that ψ is small. Now ϕ is a Lagrangian coordinate of position of the moon. If A, B, and C be the permanent parts of the principal moments of inertia of the moon about its centre, the rate of change of angular momentum of the moon about an axis through its centre perpendicular to the ecliptic is $C\ddot{\phi}$.

The couple due to the earth's attraction on the moon's permanent ellipticity of figure is $-\dfrac{3fM(B-A)}{c^3}\cos\psi\sin\psi$, where M is the mass of the earth and c its mean distance. In addition there may be a couple due to internal tidal friction. Let us, however, examine the consequences if this is absent. The equation of motion is

$$C\ddot{\phi} = -\frac{3fM(B-A)}{c^3}\cos\psi\sin\psi \qquad \text{.........(2),}$$

or, if we use (2) and neglect ψ^2,

$$\psi + 3n^2\frac{B-A}{C}\psi = -\dot{\theta} \qquad \text{.........(3).}$$

But $\dot{\theta}$ is the rate of increase of the moon's longitude, and is therefore equal to n. Thus

$$\ddot{\theta} = \frac{dn}{dt} \qquad \text{.........(4),}$$

which is very small. θ is of order $\dfrac{1}{n}\left(\dfrac{dn}{dt}\right)^2$. A sufficient approximation to a solution of (3) is therefore to be got by supposing that ψ also varies very slowly, so that $\ddot{\psi}$ can be neglected. Then

$$\psi = -\frac{C}{B-A}\frac{dn}{3n^2\,dt} \qquad \text{.........(5).}$$

Evidently ψ/ψ is of order $\left(\dfrac{1}{n}\dfrac{dn}{dt}\right)^2$, and therefore our assumption that it is negligible leads to self-consistent results. Now at present

$$\frac{dn}{dt} = -3n_0\frac{d\xi}{dt},$$

$$\kappa\Omega_0\frac{d\xi}{dt} + \frac{d\Omega}{dt} = 0 \text{ from 14·51 (6),}$$

and therefore

$$\frac{1}{3n^2}\frac{dn}{dt} = \frac{1}{\kappa n\Omega}\frac{d\Omega}{dt}$$
$$= -4\times 10^{-13}.$$

Again, $(B-A)/C$ is 0·00047. Thus

$$\psi = 10^{-9} \text{ nearly}$$
$$= 2'' \times 10^{-4}.$$

Thus even in the absence of any friction in the moon at present, the recession of the moon from the earth would produce a perfectly imperceptible deviation of the moon's longest axis from the line of centres.

It may be pointed out that, small though this deviation is, it has an important dynamical effect. The moon's longest axis pointing systematic-

ally to one side of the earth causes the couple on the moon produced by the earth's attraction on the moon's equatorial protuberance to be on an average negative. It is this couple that reduces the moon's rate of rotation and makes it remain equal to the rate of revolution.

14·62. It may be thought, however, that even though the moon's permanent ellipticity keeps the particular integral of 14·61 (3) small, the complementary functions may increase considerably; in other words, the amplitude of the moon's free libration in longitude may become great. This, however, is not the case. The period of the free libration is seen from (3) to be $\frac{2\pi}{n}\left(\frac{C}{3(B-A)}\right)^{\frac{1}{2}}$, or 27 months, and clearly remains for all time proportional to the moon's period of revolution. Thus the change in the period of oscillation during a complete oscillation is only a small fraction of the period itself. In these circumstances it is known* that a first approximation to the solution, valid for all time, of an equation

$$\frac{d^2y}{dx^2} + \chi y = 0,$$

is

$$y = A\chi^{-\frac{1}{4}} \cos \int^x \chi^{\frac{1}{2}}\, dx + B\chi^{-\frac{1}{4}} \sin \int^x \chi^{\frac{1}{2}}\, dx,$$

where A and B are arbitrary constants. Thus the amplitude is proportional to the square root of the period. If it was small at the moon's last adjustment to the hydrostatic state, which we have seen probably took place when the period of revolution was about 6·3 of our present days, it would by now have been multiplied by $(27{\cdot}3/6{\cdot}3)^{\frac{1}{2}}$ or 2·1. It would therefore still be small.

To explain the facts that the moon keeps the same face always towards the earth, and that its free libration in longitude is imperceptible, it is therefore unnecessary for tidal friction to be still operating in its interior. It is enough that tidal friction should have been sufficient to produce these conditions before, or soon after, solidification, which is highly probable; once produced, they would be permanently maintained by the earth's attraction on the moon's equatorial protuberance.

14·7. *Tidal Friction on other Planets and Satellites.* In other systems than our own, tidal friction may be expected to operate in four ways:

1. Tides raised in the satellites by their primaries will tend to make them keep the same face towards their primaries.
2. Tides raised in the primaries by the satellites will alter the rates of rotation of the primaries.
3. Tides raised in the primaries by the satellites will alter the distances of the satellites from the primaries.
4. Solar tides will affect the rates of rotation of the planets.

These four effects may be considered separately.

* Cf. a forthcoming paper by the writer in *Proc. Lond. Math. Soc.*

14·71. We notice that in 14·6 (1) m is the mass of the tide-raising body, and M and A refer to the deformed body. Thus fm/c^3, when we are considering tides raised in a satellite, is practically n^2, and M/A^3 is proportional to the density of the satellite. With the usual assumptions as to uniformity of physical constitution among satellites, it follows that the rate of change of velocity of rotation in a satellite is proportional to the fourth power of its mean motion. We should therefore expect that all satellites whose periods are less than that of the moon would turn the same faces permanently towards their primaries; satellites of longer periods may not yet have reached this state. All satellites whose rotation periods are known do actually keep the same faces towards their primaries; they include the great satellites of Jupiter, and also Iapetus. The period of the latter is nearly three months.

14·72. It can be definitely asserted that no satellite other than the moon has produced a considerable effect on the rotation of its primary. For the effect of tidal friction is to transfer angular momentum from the rotation of the primary to the revolution of the satellite, and if the rotation of any planet had been much affected in this way the angular momentum of the satellite's revolution would be comparable with that of the planet's rotation. This is true of no satellite except the moon; the orbital angular momenta of all other satellites are insignificant in comparison with the rotational angular momenta of their primaries.

This fact is consistent with the tidal lags on the great planets being comparable with that on the earth. For if a satellite have radius a and the same density as its primary,

$$\frac{m}{M}\left(\frac{A}{a}\right)^3 = 1,$$

and therefore

$$\left(\frac{m}{M}\right)^2\left(\frac{A}{c}\right)^6 = \left(\frac{a}{c}\right)^6,$$

and thus, for the same tidal lag, the rate of change of angular velocity should be proportional to the product of $\sin 2\epsilon$ into the sixth power of the apparent diameter of the satellite as seen from its primary. No other satellite subtends at the centre of its primary a greater angle than the moon, though the moon's apparent size is approached by Phobos and J I. It is therefore possible that the satellites have not affected the rotations of their primaries considerably, even if the primaries show as great tidal lags as the earth.

14·73. The argument is readily extended to the effect of the tides raised by the sun. The sun subtends at Jupiter an angle of only 6′, which is less than that subtended by J I, whose density is very similar. The lags of the tides in Jupiter raised by the sun and J I cannot be very different. Thus the effect of solar tidal friction on Jupiter must be less

than that of J I, which we already know to be insignificant. The effects on Saturn, Uranus, and Neptune must be still smaller.

Mars is probably more nearly comparable with the earth, and tidal dissipation on it may occur in shallow seas. The apparent diameter of the sun as seen from it is 21′, and that of Phobos, inferred from the brightness, is about 20′. Phobos, however, probably has a greater density than the sun, and thus the solar tides are probably less important to Mars than those raised by Phobos. They therefore can hardly have affected the rotation of the planet much.

Venus and Mercury are in a different position. Supposing the tidal lags in these planets to be equal to that in the earth, we see that the rate of change of Venus's speed of rotation must be $(0{\cdot}72)^{-6}$, or 7·2 times the effect of the sun on the earth. The corresponding ratio for Mercury is about 300. Venus has an atmosphere and may have shallow seas; Mercury, on the other hand, has none, and friction in it, if any, must be bodily. It appears probable, then, that if Venus were now rotating in the same time as the earth, its rate of angular retardation would be rather more than twice that of the earth on account of the sun and moon together; but this rate is not nearly so fast as the former rate of angular retardation of the earth when the moon was nearer. If then Venus once had a short period of rotation, of a few hours only, it is probable that it would not yet have been made to rotate in so long a period as the earth; if, however, it had a period such as the earth has now, it would have been very much lengthened by solar tidal friction. It is now known from spectroscopic observations that the rotation of Venus cannot be nearly so fast as that of the earth, so that we may infer that its original period of rotation was probably not very fast.

Mercury has probably had a history similar to that of the moon. Its periods of rotation and revolution were made equal by solar tidal friction, probably before it was thoroughly solid, and this condition has been maintained ever since by bodily friction; the distance of this planet from the sun cannot have changed much, but an ellipsoidal inequality of the planet's figure may have been produced in solidification in the way suggested for the moon, and have been afterwards maintained by the strength of the material. This may have since kept the same face of Mercury turned towards the sun, in the manner suggested for the moon in 14·61.

14·74. The rate of increase of a satellite's distance through tidal friction is given by 14·2 (3)

$$\frac{Mm}{M+m}c^2 n\frac{d\xi}{dt} = N \qquad \text{.........(1)}.$$

From 14·6 (1)

$$\frac{N}{C} = \tfrac{18}{5}\pi f\rho \sin 2\epsilon \left(\frac{m}{M}\right)^2 \left(\frac{A}{c}\right)^6 \qquad \text{.........(2)},$$

and we know

$$C = \tfrac{1}{3}MA^2 \qquad \text{.........(3)}.$$

On eliminating N and C we find

$$\frac{d\xi}{dt} = \tfrac{6}{5}\pi f\rho \sin 2\epsilon \, \frac{(M+m)\,m}{M^2 n}\left(\frac{A}{c}\right)^8$$

$$= \tfrac{6}{5}\pi f\rho \sin 2\epsilon \, \frac{m}{Mn}\left(\frac{A}{c}\right)^8 \qquad \ldots\ldots\ldots(4),$$

if we treat $(M+m)/M$ as equal to unity.

For J I, m/M is 1/20,000, as against 1/80 for the moon; n has about 20 times its value for the moon; A/c has about 10 times its value for the earth and moon. Thus $d\xi/dt$ has about 2×10^4 times the value corresponding to the moon at present. If then the constitution of Jupiter resembled that of the earth, the mean distance of J I would be doubled in about 6×10^5 years.

For Titan, again, $d\xi/dt$ should be about 60 times what it is for the moon: and for the satellite of Neptune it should be about 2000 times as great.

It appears then that unless the great planets approximate very closely to perfect elasticity, giving much smaller tidal lags than the average in our ocean, the evolution of their nearer large satellites must have been dominated by tidal friction. This will not be true of the more remote satellites, since the factor $(A/c)^8$ diminishes very rapidly as c increases.

Phobos presents a further difficulty. It is readily seen that, on the same physical hypotheses, $d\xi/dt$ for it should be 8000 times as great as for the moon. This result is even more embarrassing than that for J I; for Phobos revolves more rapidly than Mars rotates, and therefore tidal friction, in accordance with the argument of 14·52, will make Phobos approach the planet instead of receding from it. Thus Phobos would have been precipitated on the surface of the planet if it was as old as the moon.

It is possible, however, that our estimate will have to be much reduced, for two reasons. Mars may have little fluid on its surface, and therefore no dissipation in its shallow seas. If so, tidal friction within it must be bodily. A reduction of $d\xi/dt$ to a small fraction of the value just found may therefore have to be made. A further reduction will be needed on account of the small size of Mars. In a small planet the elastic tide is a small fraction of the hydrostatic tide; thus, in Mars the height of the tide may well be reduced to $\frac{1}{20}$ of its hydrostatic value by rigidity. For these reasons it is quite possible that the evolution of the orbit of Phobos may proceed no faster than that of the moon.

We notice, however, that the arguments just used for Mars are applicable with greater force to Mercury. Thus if the solar tides are supposed to have made the latter planet keep the same face turned to the sun, we seem to be driven to adopt the theory of 14·61 for Mercury as well as the moon.

14·8. *Summary.* Friction, either in oceanic or in bodily tides, must produce a continual diminution in the rate of rotation of the earth, and

continual increases in the mean distances of the moon and sun from the earth. These changes together make the moon and sun appear to have, relative to the stars, slow secular accelerations which can be found by comparison of modern observations with ancient ones of eclipses and occultations. The secular acceleration of the sun found in this way is a somewhat larger fraction of that of the moon than would theoretically have been expected. The rate of dissipation of energy required to explain these accelerations is about $1{\cdot}4 \times 10^{19}$ ergs per second on an average.

It is found that tidal friction in the open ocean cannot account for more than about a thousandth of this, but that the greater part of it can be explained quantitatively by tidal friction in shallow seas. There is no reason to believe these seas unable to account for the whole of it, and therefore tidal friction in the body of the earth may be inappreciable. It appears, indeed, that if a considerable fraction of the secular accelerations were due to elastic afterworking, earthquake waves could not be transmitted; and that if it were due to plasticity in a homogeneous earth, the 14-monthly variation of latitude could not exist.

By extrapolation it is found that the period of the earth's rotation may have changed by about 4 hours during geological time, and that the geological effects of a change of rotation may possibly be appreciable. Tidal friction is capable of explaining how the moon came to recede from close proximity to the earth to its present distance; the time required does not appear likely to be prohibitive. The moon will ultimately recede till its period of revolution and the period of the earth's rotation are both equal to about 47 of our present days. When this takes place the moon will gradually approach the earth again, ultimately passing within Roche's limit and being broken up. (It may be remarked that on the resonance theory, on account of the great extension of the earth at the time, the moon was outside Roche's limit when it was first formed.)

Tidal friction readily accounts for the fact that the moon always keeps the same face towards the earth; it is sufficient that this condition should have been brought about before the moon solidified, for afterwards it could have been maintained by purely gravitational causes. In the same way, Mercury has probably been made to keep the same face towards the sun. This result is probably applicable to all satellites whose periods are less than that of the moon. No planet except the earth has had its rotation much affected by tides raised by a satellite. It is difficult to make any inference about Venus, except that its rotation period has not been lengthened to such an extent as that of the earth. The orbits of some satellites may have been appreciably affected by tidal friction, notably J I and Phobos, but further inferences cannot be made without more knowledge about the physical conditions on their primaries.

CHAPTER XV

The Variation of Latitude

"What is the use of the axis of the earth?" — School examination paper.
"Parturiunt montes, nascetur ridiculus mus." — HORACE, *Ars Poetica.*

15·1. The latitude of a given observatory was found by Chandler in 1891 to be subject to a small variation, consisting apparently of two parts, one with a period of a year, and the other about 14 months. Both had amplitudes of the order of a tenth of a second of arc.

Now the latitude of a station is by definition the mean of the altitudes of a circumpolar star when it crosses the meridian above and below the pole, refraction being supposed absent*. The altitude is the complement of the zenith distance, and the zenith is in the direction exactly opposite to gravity; and the mean of the altitudes of the star at its two passages is evidently the altitude of the celestial pole, which is itself in the direction of the earth's axis of rotation. Thus the colatitude of a station is the angle between local gravity and the earth's instantaneous axis of rotation. A change in it must therefore be attributed to a change in the direction either of local gravity or of the earth's instantaneous axis of rotation. The latter explanation is the one generally adopted. One reason for this is that a periodic motion of the earth's instantaneous axis is to be expected on theoretical grounds, having been predicted by Euler. Again, the displacement of the pole can be specified by using only two coordinates, and when these are given all the changes of latitude produced by it can be inferred. The change will evidently be the same for all stations with the same longitude. Thus observations of the changes of latitude of every station (after the first two) give an independent test of the hypothesis that all the changes are due to displacements of the axis of rotation. It is actually found that observations at different stations are consistent with the hypothesis that the changes are produced in this way; the hypothesis therefore has a high probability.

The polar motion required to account for several years' observations at different stations was found to be composed of two parts. The part with a period of a year is elliptical The remainder is very irregular, but its major features can be represented by a circular motion in a period of 14 months, the amplitude and epoch being subject to slow but continual variation. The whole motion of the pole over the earth's surface is usually referred to as the 'variation of latitude,' though this term is better reserved

* Refraction is great at the lower passage, and makes this method of finding the latitude inaccurate in practice; the practical method is that of Talcott, depending on observations of different stars near the zenith, and is discussed in *Publications of R. Observatory, Greenwich,* 'Observations with the Cookson floating telescope, 1911–18.'

for the actual changes of latitude at the observing stations, rather than the polar motion inferred from them.

15·2. The theories of the annual and 14-monthly motions may be treated together up to a certain point. Let a system of axes at the centre of the earth be defined as follows:

The axis of x is in the plane of the equator, and intersects the meridian of Greenwich.

The axis of y is in the plane of the equator, and is 90° east of Greenwich.

The axis of z points to the mean position of the north pole.

Let the instantaneous angular velocities of these axes, referred to inertial axes instantaneously occupying the same positions, be θ_1, θ_2, θ_3. Let n be the resultant angular velocity of the earth. Then the direction cosines of the instantaneous axis are $(\theta_1/n, \theta_2/n, \theta_3/n)$.

It was seen in our discussion of the figure of the earth that the material of the interior of the earth probably behaves almost as a fluid to stresses of long duration, but the behaviour of earthquake waves shows that most of it behaves as an elastic solid when the period of the stress is a few seconds. It remains to be seen how it behaves when the period of the disturbance is 14 months or a year. We may expect that when the axis of rotation shifts, the shift will produce an elastic deformation of the earth. The actual ellipticity of the earth may be considered as composed of two parts, one of which is unaffected by small disturbances of short period, while the other is due to the elastic strain produced by rotation, and would disappear if the rotation were removed and no permanent set took place. Provided that during the phenomenon of the variation of latitude no appreciable permanent set occurs, it will be legitimate to treat these two parts separately, the first keeping throughout the motion fixed with regard to the axes of reference, and the second remaining symmetrical about the instantaneous axis.

Let I be the portion of the difference between the moments of inertia of the earth about the polar axis and an axis perpendicular to it that is due to elastic strain. Let the products of inertia be F, G, H. Suppose that if the axis of rotation coincided with the axis of z the moments of inertia would be A', B', $C' + I$. Then the angular momentum of the earth is composed of three parts:

(1) The part due to the motion as if the earth were a rigid body, rigidly attached to the axes of reference, and having moments and products of inertia $(A'B'C'FGH)$. This has components

$$A'\theta_1 - H\theta_2 - G\theta_3,\quad B'\theta_2 - F\theta_3 - H\theta_1,\quad C'\theta_3 - G\theta_1 - F\theta_2.$$

(2) The part due to rotational strain. This gives an angular momentum In about the instantaneous axis, and therefore has components $I\theta_1$, $I\theta_2$, $I\theta_3$ about the axes of reference.

(3) There is a further part due to any motion relative to the axes not

produced by rotation. In simple deformation produced by vertical movement the velocities involved in these displacements would necessarily have no moment about the centre of the earth. If matter was collecting at a point, the velocities of approach to the point being symmetrical about it, the moments would likewise be zero. In all ordinary displacements it appears* that this part is either zero or very small in comparison with $A'\theta_1$. If it should subsequently be found that there are large periodic motions relative to the earth as a whole that do produce a considerable angular momentum, then these will have to be included in a more detailed treatment. They are not considered here.

If, then, the total angular momenta are (h_1, h_2, h_3), we have

$$\left.\begin{aligned} h_1 &= (A' + I)\,\theta_1 - H\theta_2 - G\theta_3 \\ h_2 &= (B' + I)\,\theta_2 - F\theta_3 - H\theta_1 \\ h_3 &= (C' + I)\,\theta_3 - G\theta_1 - F\theta_2 \end{aligned}\right\} \quad \text{.........(1).}$$

The equations of rotational motion, in the absence of external forces, are therefore

$$\left.\begin{aligned} &\frac{d}{dt}\{(A' + I)\,\theta_1 - H\theta_2 - G\theta_3\} - \{(B' + I)\,\theta_2 - F\theta_3 - H\theta_1\}\,\theta_3 \\ &\qquad\qquad + \{(C' + I)\,\theta_3 - G\theta_1 - F\theta_2\}\,\theta_2 = 0 \\ &\frac{d}{dt}\{(B' + I)\,\theta_2 - F\theta_3 - H\theta_1\} - \{(C' + I)\,\theta_3 - G\theta_1 - F\theta_2\}\,\theta_1 \\ &\qquad\qquad + \{(A' + I)\,\theta_1 - H\theta_2 - G\theta_3\}\,\theta_3 = 0 \\ &\frac{d}{dt}\{(C' + I)\,\theta_3 - G\theta_1 - F\theta_2\} - \{(A' + I)\,\theta_1 - H\theta_2 - G\theta_3\}\,\theta_2 \\ &\qquad\qquad + \{(B' + I)\,\theta_2 - F\theta_3 - H\theta_1\}\,\theta_1 = 0 \end{aligned}\right\} \quad \text{......(2).}$$

In the undisturbed state the earth may be regarded as dynamically symmetrical about the polar axis. Let the undisturbed values of A', B', C', be A, A, C_0. Put

$$A' = A + A_1;\; B' = A + B_1;\; C' = C_0 + C_1 \quad \text{.........(3).}$$

Then A_1, B_1, C_1, F, G, H, θ_1, θ_2 are all small quantities of the first order. If we neglect their squares and products, the third equation of motion reduces to

$$\frac{d}{dt}\{(C' + I)\,\theta_3\} = 0 \quad \text{.........(4),}$$

and therefore $(C' + I)\,\theta_3$ is a constant, equal to $(C_0 + I)\,n$, the angular momentum due to rotation.

The first two equations reduce to

$$(A + I)\frac{d\theta_1}{dt} - (C_0 + I)\,n\,\frac{d}{dt}\left(\frac{G}{C_0 + I}\right) - (A - C_0)\,n\theta_2 + Fn^2 = 0,$$

$$(A + I)\frac{d\theta_2}{dt} - (C_0 + I)\,n\,\frac{d}{dt}\left(\frac{F}{C_0 + I}\right) - (C_0 - A)\,n\theta_1 - Gn^2 = 0,$$

* This point has been considered in greater detail by Sir G. H. Darwin, *Scientific Papers*, 3, 40.

which may be written

$$\left.\begin{aligned}\frac{d\theta_1}{dt}+\frac{C_0-A}{A+I}n\theta_2&=\frac{n}{A+I}\frac{dG}{dt}-\frac{Fn^2}{A+I}\\ \frac{d\theta_2}{dt}-\frac{C_0-A}{A+I}n\theta_1&=\frac{n}{A+I}\frac{dF}{dt}+\frac{Gn^2}{A+I}\end{aligned}\right\} \qquad \text{.........(5).}$$

These equations determine the polar motion. Evidently the terms on the right involving d/dt are smaller than the others in a ratio of the order of 1/400 for annual or 14-monthly motions, and may be neglected. The direction cosines of the instantaneous axis being denoted by $(l, m, 1)$, we can write (5) in the form

$$\left.\begin{aligned}\frac{dl}{dt}+\frac{n}{\tau}m&=-\frac{Fn}{A+I}\\ \frac{dm}{dt}-\frac{n}{\tau}l&=\frac{Gn}{A+I}\end{aligned}\right\} \qquad \text{.........(6),}$$

where

$$\tau=\frac{A+I}{C_0-A} \qquad \text{.........(7).}$$

15·3. Let us now consider the complementary functions. Multiplying the second of (6) by ι, where $\iota^2=-1$, adding it to the first, and writing

$$l+\iota m=w \qquad \text{.........(1),}$$

we see that the complementary functions satisfy

$$\frac{dw}{dt}-\frac{\iota n}{\tau}w=0 \qquad \text{.........(2),}$$

whence

$$w=\alpha e^{\iota n(t-t_0)/\tau} \qquad \text{.........(3),}$$

where α and t_0 are real arbitrary constants. Then

$$l=\alpha\cos n(t-t_0)/\tau;\quad m=\alpha\sin n(t-t_0)/\tau \qquad \text{.........(4).}$$

This represents a motion in which the instantaneous axis remains at a constant angle α to the axis of z, moving about it with constant angular velocity n/τ from west to east. Thus the period of the motion is τ sidereal days.

It is natural to identify this free motion with the 14-monthly period. Its period then provides a determination of the important constant τ. We see that this is practically equal to $A/(C_0-A)$, since C_0-A and I are both small fractions of C_0. It is therefore nearly the reciprocal of the part of the precessional constant that arises from the permanent oblateness. A physical reason for this is readily seen. In a perfectly rigid body whose total moments of inertia were A and C_0, τ would be $A/(C_0-A)$. Thus our result expresses the obvious fact that the part of the ellipticity that keeps perfectly symmetrical about the axis of rotation can have no effect in displacing that axis.

If now C is the total moment of inertia of the earth about its polar axis, so that

$$C=C_0+I, \qquad \text{.........(5),}$$

the precessional constant is $(C - A)/C$. The part of the precessional constant arising from purely elastic strain is therefore $\dfrac{C-A}{C} - \dfrac{1}{\tau}$. This affords an important datum about the elasticity of the earth as a whole.

We notice that if the earth were perfectly rigid and had its actual moments of inertia, τ and $A/(C - A)$ would be equal. Thus the period of the vibration would be 305 days. This result was found by Euler, and the motion is therefore often called the 'Eulerian nutation.' If on the other hand the earth had no rigidity, the whole of the actual oblateness would be due to strain; the oblateness would always be exactly symmetrical about the axis of rotation, $C - A$ would be zero, and the period would be infinite. This may be interpreted by returning to 15·2 (6). This gives now

$$\frac{dl}{dt} = \frac{dm}{dt} = 0,$$

so that the direction of the axis is constant. In other words, a fluid earth would settle down into a motion with the axis of rotation in a fixed direction and the whole mass symmetrical about it; no Eulerian nutation would be possible. Actually τ is about 430, subject to an uncertainty of a few units.

Historically, Euler's prediction caused astronomers to look for an oscillation with a period of ten months, and it was only after repeated failures to find such a periodicity that the 14-monthly period was discovered. After its discovery, Newcomb pointed out that the lengthening of the period could be attributed to elasticity; further developments of the theory are due to Hough, Love, and Larmor.

15·4. We can now proceed to consider the forced motion. It will be convenient to introduce an auxiliary axis called the 'axis of inertia,' whose extremity is the 'pole of inertia.' If λ, μ, ν be the direction cosines of any diameter whatever of the earth, the moment of inertia about it arising from the system of moments and products $A'B'C'FGH$ is

$$J = A'\lambda^2 + B'\mu^2 + C'\nu^2 - 2F\mu\nu - 2G\nu\lambda - 2H\lambda\mu \quad \text{.........(1)},$$

where
$$\lambda^2 + \mu^2 + \nu^2 = 1 \quad \text{.........(2)}.$$

The function J will be stationary, subject to (2), if λ, μ and ν are such that

$$\frac{\partial J}{\partial \lambda} d\lambda + \frac{\partial J}{\partial \mu} d\mu + \frac{\partial J}{\partial \nu} d\nu,$$

and
$$\lambda d\lambda + \mu d\mu + \nu d\nu$$

vanish simultaneously for all values of the ratios of $d\lambda$, $d\mu$, $d\nu$. Then there will be a quantity σ such that

$$\frac{\partial J}{\partial \lambda} = \sigma\lambda; \quad \frac{\partial J}{\partial \mu} = \sigma\mu; \quad \frac{\partial J}{\partial \nu} = \sigma\nu \quad \text{.........(3)}.$$

These give

$$\left.\begin{aligned} (A' - \sigma)\lambda - H\mu - G\nu &= 0 \\ -H\lambda + (B' - \sigma)\mu - F\nu &= 0 \\ -G\lambda - F\mu + (C' - \sigma)\nu &= 0 \end{aligned}\right\} \quad \text{.........(4)}.$$

There must evidently be an axis of maximum moment of inertia near the undisturbed one, and therefore near the z axis. Hence for this axis λ and μ will be small, and ν will differ from 1 only by small quantities of the second order. Then (4) become

$$\left.\begin{aligned}(A-\sigma)\lambda - G &= 0\\ (A-\sigma)\mu - F &= 0\\ (C_0-\sigma) \qquad &= 0\end{aligned}\right\} \qquad \text{.........(5),}$$

and therefore

$$\left.\begin{aligned}\lambda &= -\frac{G}{C_0-A} = -\frac{G\tau}{A}\\ \mu &= -\frac{F}{C_0-A} = -\frac{F\tau}{A}\end{aligned}\right\} \qquad \text{.........(6).}$$

This axis of maximum moment of inertia may be called the axis of inertia. Its utility is that it makes it possible to reduce 15·2 (6) to the simple form

$$\left.\begin{aligned}\frac{\tau}{n}\frac{dl}{dt} + m &= \mu\\ \frac{\tau}{n}\frac{dm}{dt} - l &= -\lambda\end{aligned}\right\} \qquad \text{.........(7).}$$

We notice that the effect of any accumulation of matter on the earth's surface is to displace the axis of inertia away from the added matter.

Now l and m may be found from the polar motion, being indeed the component angular displacements of the pole. If then some explanation of the annual variation of latitude is suggested, and we can find the corresponding annual variation of F and G, we can compare the two sides of equations (7) and obtain a quantitative test of the hypothesis.

15·5. We have now to express F and G, and hence λ and μ, in terms of the mass-transference that occurs during the year. Suppose the annual component of the mass per unit area on an element of the earth's surface to be $\frac{1}{2}\sigma\cos\odot + \frac{1}{2}\sigma'\sin\odot$, where $\odot$ is the sun's longitude. Then

$$\begin{aligned}F &= \iiint \rho yz\,dx\,dy\,dz \text{ through the volume}\\ &= \iint (\tfrac{1}{2}\sigma\cos\odot + \tfrac{1}{2}\sigma'\sin\odot)\, yz\,dS \text{ over the surface}\\ &= a^4\int_0^\pi\int_0^{2\pi}(\tfrac{1}{2}\sigma\cos\odot + \tfrac{1}{2}\sigma'\sin\odot)\sin^2\theta\cos\theta\sin\phi\,d\theta\,d\phi \quad \text{...(1),}\\ G &= a^4\int_0^\pi\int_0^{2\pi}(\tfrac{1}{2}\sigma\cos\odot + \tfrac{1}{2}\sigma'\sin\odot)\sin^2\theta\cos\theta\cos\phi\,d\theta\,d\phi \quad \text{...(2),}\end{aligned}$$

where a is the radius of the earth, θ the colatitude, and ϕ the longitude east of Greenwich. σ and σ' are functions of θ and ϕ.

These values of F and G, however, require to be corrected for the effect of elastic yielding in the earth under the pressure of the superficial matter. If the variable part of the mass per unit area is m, let m be supposed

expanded in spherical harmonics. Let one term in this expansion be $\alpha \sin\theta \cos\theta \sin\phi$, $= \alpha S_2$, say. Then F has exactly the same value as if m were actually equal to αS_2. Now the strain due to any harmonic in m can be treated separately, and the radial displacement will be proportional to the same harmonic as the surface density. Hence no harmonic other than S_2 can contribute anything to F. All we need to find, then, is the strain in the earth due to a surface layer of density αS_2. Unfortunately our knowledge of the elastic properties of the earth is not adequate to find this; all we can say is that the strain will reduce F and G in a constant ratio, which we shall call κ. If the elastic constants had the values inferred from earthquakes, κ would be about $\frac{3}{4}$; but the elastic constants appropriate to forces with periods such as we are here considering appear to be much less than those appropriate to earthquake waves.

Using $A = \frac{1}{3}Ma^2$, where M is the mass of the earth, we have

$$\lambda = \kappa\lambda_0; \quad \mu = \kappa\mu_0; \qquad \text{.........(3)},$$

where

$$\left.\begin{aligned}\lambda_0 &= -9''{\cdot}1 \times 10^{-3}\int_0^{\pi}\int_0^{2\pi}(\sigma\cos\odot + \sigma'\sin\odot)\sin^2\theta\cos\theta\cos\phi\, d\theta\, d\phi\\ \mu_0 &= -9''{\cdot}1 \times 10^{-3}\int_0^{\pi}\int_0^{2\pi}(\sigma\cos\odot + \sigma'\sin\odot)\sin^2\theta\cos\theta\sin\phi\, d\theta\, d\phi\end{aligned}\right\}\text{...(4)},$$

where σ and σ' must now be expressed in grams per square centimetre.

We are now in a position to treat various putative causes separately. There are several obvious annual changes in the distribution of mass over the surface that could affect the products of inertia. Evidently, if the whole of the surface of the earth were of similar character, either land or sea, or even if the distribution of land and sea were symmetrical about the polar axis, F and G would necessarily be zero, and there could be no displacement of the axis of rotation. The chief methods of redistribution of matter over the surface in the course of a year are apparently

(1) the seasonal variation in the distribution of air over the surface, shown by the variation of atmospheric pressure observed at the surface;

(2) precipitation of snow, which accumulates in places throughout the winter;

(3) periodical changes in vegetation, such as the formation of deciduous parts of trees, the rise of sap in trees, and the formation of annual parts of herbs.

15·51. The numerical data for the estimation of the effects of the annual redistribution of air have been found from the annual variation of pressure over the earth, given in Bartholomew's *Meteorological Atlas*; though a recalculation with more modern data may be desirable. Allowance must be made for the effect of the pressure on the sea surface in redistributing the water of the ocean. If m be taken to refer to the mass

of air and water together per unit surface, we have from the law of indestructibility of matter $\iint m ds = 0$ ....(1)

taken over the whole surface of the earth, both land and water.

But for an annual motion the elevation of the ocean surface must have practically its equilibrium value. In other words, the water will adjust itself so that the pressure is uniform over any level surface and therefore has no tendency to produce horizontal movement. In these circumstances the mass per unit area of the surface above a given level must be uniform over the ocean; and the periodical variation of the mass per unit area is therefore the same all over the ocean. Let its value be m'. Then

$$m' \iint dS + \iint m dS = 0 \qquad \text{.........(2)}$$

by (1); where the first integral extends over the sea and the second over the land. If, then, m_1 be the mean value of m over the land,

$$\begin{aligned} m' &= -m_1 (\text{area of land})/(\text{area of sea}) \\ &= -\cdot 40 m_1 \qquad \text{.........(3).} \end{aligned}$$

Next,

$$\begin{aligned} F &= \iint myzdS \text{ over the whole surface} \\ &= \iint m'yzdS \text{ over the ocean} \\ &\quad + \iint myzdS \text{ over the land} \\ &= \iint (m + 0{\cdot}40m_1)\, yzdS \text{ over the land} \qquad \text{.........(4).} \end{aligned}$$

Use has here been made of the fact that $\iint yzdS$, taken over the whole surface, is zero. Similarly

$$G = \iint (m + 0{\cdot}40m_1)\, zxdS \text{ over the land} \qquad \text{.........(5).}$$

On carrying out the numerical integrations, we find

$$\left.\begin{aligned} \lambda_0 &= 0''{\cdot}0040 \cos \odot + 0''{\cdot}0051 \sin \odot \\ \mu_0 &= -0''{\cdot}0134 \cos \odot + 0''{\cdot}0659 \sin \odot \end{aligned}\right\} \qquad \text{.........(6).}$$

A correction is, however, required; or rather, a correction already applied needs to be removed. The pressures given in the meteorological charts are not the true atmospheric pressures at the places concerned, but these pressures reduced to sea-level according to a formula given by Laplace* which depends in part upon the temperature. To find the mass of air over a given area, however, we need, not these modified values, but the true local pressures; the correction included in the charts therefore needs to be taken out again. This requires that to the values of λ_0 and μ_0 just found we must add†

$$\left.\begin{aligned} \lambda_0 &= 0''{\cdot}0000 \cos \odot - 0''{\cdot}0014 \sin \odot \\ \mu_0 &= 0''{\cdot}0128 \cos \odot - 0''{\cdot}0315 \sin \odot \end{aligned}\right\} \qquad \text{.........(7).}$$

The atmosphere and ocean together therefore contribute an amount

$$\left.\begin{aligned} \lambda_0 &= 0''{\cdot}0040 \cos \odot + 0''{\cdot}0037 \sin \odot \\ \mu_0 &= -0''{\cdot}0006 \cos \odot + 0''{\cdot}0344 \sin \odot \end{aligned}\right\} \qquad \text{.........(8).}$$

* *Mécanique Céleste*, edition of 1880, 4, 294.

† The numerical data for this computation and that of (6) are in *M.N.R.A.S.* 76, 1916, 499–525.

The greater part of the contribution comes from the seasonal change in Central Asia, whose chief meteorological correlate is the monsoons. For this reason the greatest part of the displacement of the pole of inertia is towards Central Asia in summer, when the amount of air there is a minimum, and away from it in winter.

It may be remarked that it would be incorrect to say that the part of the annual variation of latitude derived from (8) is due to the annual variation of pressure. It has been seen that the annual variation of latitude depends simply on the products of inertia of the earth as a whole, the atmosphere being included, and the atmospheric pressure is used only as an observable quantity that enables us to infer the mass per unit surface; that it does not affect the motion directly is seen from the facts that it is equal to the product of gravity into the mass per unit area, and that gravity nowhere occurs in the results.

15·52. *The Effect of Precipitation.* In this investigation we are concerned only with those modes of transport of matter that cause a variation in the mass per unit area over the surface. For instance, if rain falls upon a particular spot, and at once runs away to the sea, it contributes nothing to the elements of the products of inertia corresponding to that point. Similarly, if it evaporates at once, it contributes nothing to the elements derived from the solid and liquid parts of the earth; the presence of the water vapour in the atmosphere, however, will affect the mass of the atmosphere locally, and consequently its mass will be included in the variations already inferred from the distribution of atmospheric pressure. Thus the effect of water vapour in the atmosphere has already been considered. This section is concerned only with those parts of the precipitation that remain on the ground for a considerable time. These are, first, the water that is absorbed by the soil; second, the snow that lies on the ground in winter. Data for the first part are almost entirely lacking. In this country the soil is wetter in winter than in summer, principally owing to the reduced evaporation; the annual variation of rainfall is not nearly so great as that of soil-moisture. In countries, however, where most of the rain falls in summer, this state of affairs may be reversed. On the whole it seems probable that the effect of soil-moisture is small in comparison with that of snowfall.

A rough estimate of the effect of snowfall can be made easily. Snow accumulates in winter to a considerable depth over a great part of Canada and the United States, and over the greater part of Asia down to about the latitude of the Tian Shan. In South America the amount of snowfall is small, and in Australia it is negligible. On the assumption that in the areas of Asia and North America where snow accumulates, it does so at a uniform rate until the spring thaw, I estimate* that the contributions

* *Loc. cit.* p. 520.

to λ_0 and μ_0 are
$$\lambda_0 = -0''{\cdot}0238 \cos \odot + 0''{\cdot}0175 \sin \odot,$$
$$\mu_0 = -0''{\cdot}0126 \cos \odot + 0''{\cdot}0093 \sin \odot.$$

15·53. The effect of vegetation is small compared with that of the annual displacements of air. In woodland and natural grassland the mass of vegetation per unit area in summer exceeds that in winter by one or two grams per square centimetre, whereas the corresponding difference for atmospheric motions in Central Asia is about 10 grams per sq. cm. Thus the effect of vegetation is much less than that of atmospheric displacements; but it is sufficiently great to be observable, were it not associated with other larger displacements of the same period. The contributions from this source are roughly
$$\lambda_0 = 0''{\cdot}0017 \cos \odot - 0''{\cdot}0038 \sin \odot,$$
$$\mu_0 = 0''{\cdot}0024 \cos \odot - 0''{\cdot}0055 \sin \odot.$$

15·54. Accumulation of ice and snow round the poles may be summed up under two types.

The first includes the deposition of snow during the winter, with its removal partly by evaporation and partly by glacier flow during the summer. This part would correctly be included under snowfall. On account, however, of the symmetry of the permanently frozen areas about the poles, it seems very unlikely that this portion can contribute anything appreciable to the movement of the instantaneous axis.

The second type is due to the freezing of the oceans. Freezing in mid-ocean would necessarily lead to no change in the mass per unit area; for the ice must float in such a way that its mass is equal to that of the water displaced, and it is also necessarily equal to that of the water it was formed from. Thus the water displaced has the same volume as the frozen water, and the freezing produces no vertical movements of water such as are necessary to start horizontal motion. Ice attached to the margin of a continent may behave differently, but even in this case, on account of symmetry about the polar axis, the contribution to the motion of the axis must be small.

15·55. Summing up our results, we have the following:

	λ_0	μ_0
Atmosphere and ocean	$0''{\cdot}0037 \sin \odot + 0''{\cdot}0040 \cos \odot$	$0''{\cdot}0344 \sin \odot - 0''{\cdot}0006 \cos \odot$
Snowfall	$0''{\cdot}0175 \sin \odot - 0''{\cdot}0238 \cos \odot$	$0''{\cdot}0093 \sin \odot - 0''{\cdot}0126 \cos \odot$
Vegetation	$-0''{\cdot}0038 \sin \odot + 0''{\cdot}0017 \cos \odot$	$-0''{\cdot}0055 \sin \odot + 0''{\cdot}0024 \cos \odot$
Total	$0''{\cdot}0174 \sin \odot - 0''{\cdot}0181 \cos \odot$	$0''{\cdot}0382 \sin \odot - 0''{\cdot}0108 \cos \odot$

The actual motion of the pole of rotation was found by Kimura* to be given by
$$\left.\begin{aligned} l &= 0''{\cdot}011 \sin \odot - 0''{\cdot}096 \cos \odot \\ m &= -0''{\cdot}051 \sin \odot - 0''{\cdot}004 \cos \odot \end{aligned}\right\} \text{from } 1893{\cdot}8 \text{ to } 1899{\cdot}8,$$

* *Ast. Nach.* **181**, 4344. His x is l, and his y is $-m$.

and $\left.\begin{aligned} l &= 0''{\cdot}008 \sin \odot - 0''{\cdot}064 \cos \odot \\ m &= -0''{\cdot}056 \sin \odot + 0''{\cdot}006 \cos \odot \end{aligned}\right\}$ from 1900·0 to 1907·0,

and by the present writer* to be

$$\left.\begin{aligned} l &= 0''{\cdot}038 \sin \odot - 0''{\cdot}055 \cos \odot \\ m &= -0''{\cdot}061 \sin \odot - 0''{\cdot}017 \cos \odot \end{aligned}\right\} \text{ from 1907·1 to 1914·0.}$$

Taking $\tau = 430$, and using 15·4 (7), we derive from these

$$\left.\begin{aligned} \lambda &= 0''{\cdot}006 \sin \odot - 0''{\cdot}036 \cos \odot \\ \mu &= 0''{\cdot}062 \sin \odot + 0''{\cdot}009 \cos \odot \end{aligned}\right\} \text{ from 1893·8 to 1899·8,}$$

$$\left.\begin{aligned} \lambda &= 0''{\cdot}015 \sin \odot + 0''{\cdot}002 \cos \odot \\ \mu &= 0''{\cdot}019 \sin \odot + 0''{\cdot}015 \cos \odot \end{aligned}\right\} \text{ from 1900·0 to 1907·0,}$$

$$\left.\begin{aligned} \lambda &= 0''{\cdot}018 \sin \odot - 0''{\cdot}017 \cos \odot \\ \mu &= 0''{\cdot}004 \sin \odot + 0''{\cdot}028 \cos \odot \end{aligned}\right\} \text{ from 1907·1 to 1914·0.}$$

The most striking feature of these motions of the pole of inertia inferred from the motion of the pole of rotation is their variability among themselves. They resemble one another and the calculated motion in that all are of the same order of magnitude and retrograde (i.e. from east to west), and the largest term in the motion inferred from meteorological data, that in μ_0 involving sin $\odot$, is twice represented by the largest term in the values inferred from astronomy; but that is all that can be said. The variation is probably too great to be attributed to observational error, and certainly too great to be attributed to real variations in meteorological conditions from year to year. It is more probably due to variations in the amplitude of the 14-monthly motion: for if this changed during a 7-year interval, the 14-monthly motion could not be properly eliminated.

15·6. Returning to the question of the bodily viscosity of the earth, discussed in the last chapter, we can examine the effects of such viscosity, if it exists, on the Eulerian nutation. We saw in 14·424 that if the whole secular acceleration of the moon were to be attributed to elasticoviscosity, t_1 would have to be about 2 days, and if it were all to be attributed to firmoviscosity, t_2 would be about 270 secs. The former value would imply that the earth must behave as a fluid in any motion with a period of 14 months; thus the Eulerian nutation could not exist. The discovery of the importance of tidal friction in shallow seas in producing the secular acceleration has however shown that t_1 must be much more than 2 days. The fact that the earth's effective rigidity, determined from the observed period of latitude variation, is decidedly smaller than that inferred from seismological data, strongly suggests that considerable regions within the earth do behave as if fluid to forces of such periods. If such regions were surrounded by perfectly elastic material, without much intervening matter of intermediate viscosity, they might reduce the effective rigidity of the earth as a whole under stresses of such periods without producing much damping.

* *Loc. cit.*, pp. 523–5.

Firmoviscosity would evidently have no appreciable effect on motions with such periods. The same applies to turbulence in the ocean. The variability of the amplitude of the 14-monthly motion is, however, consistent with the existence of sufficient internal plasticity to damp it down in a few years, the motion being regenerated from other sources.

From what has been said, it will be evident that the problem of the variation of latitude remains only half solved; the explanations of the annual portion account only roughly for its character and amplitude, and there is so far no explanation of the source of the energy that maintains the 14-monthly period in spite of the damping that must exist.

APPENDIX A

The Planetesimal Hypothesis

"The man who makes no mistakes does not usually make anything."
EDWARD J. PHELPS.

A·1. The Planetesimal Hypothesis was historically the parent of the Tidal Theory of the Origin of the Solar System, elaborated in Chapter II. It was invented by T. C. Chamberlin and F. R. Moulton in the early years of the present century, and detailed accounts of it may be found in Chamberlin and Salisbury's *Textbook of Geology*, Moulton's *Introduction to Astronomy*, and Chamberlin's *The Origin of the Earth.* Like the theory here adopted, it supposes the sun to have been broken up by a passing star, and there is a general resemblance between the modes of formation and rupture of the filament on the two theories. The authors of the Planetesimal Theory suppose that two filaments were formed, projecting from the sun at diametrically opposite points. Jeans has shown that it is possible that only one filament was formed; the theory here developed works just as well with only one, and it has been suggested that it is possible that the shorter was wholly reabsorbed into the sun, even if it was ever formed. At this stage, however, the differences between the theories begin to become serious. The authors of the Planetesimal Theory believe that the planets would cool principally by adiabatic expansion, whereas it is shown here that at any rate the larger ones would cool principally by radiation from the surface. They assert further that all the planets, large and small, would form liquid drops at once, that these would quickly solidify, and that the planets formed by the aggregation of the solid particles would be themselves solid from the start. It has been shown here that, in whatever way the planets cooled, they would always pass through a liquid stage.

A·11. *Impossibility of Great Accretion.* Perhaps the most serious divergence between the two theories is, however, in the nature of the postulated resisting medium. In the present work it is supposed to be a gas, probably chiefly hydrogen; in the Planetesimal Hypothesis it is supposed to be composed of particles that solidified during the condensation of the planets, but acquired velocities so great that gravity could not retain them. These would then revolve around the sun as independent bodies; they are the 'planetesimals' that give the theory its name. It is supposed that they were afterwards largely swept up by the planets, and that their effect was to reduce the eccentricities of the planetary orbits to their present values. Now it is possible, and indeed almost certain, that many such small solid particles were actually set in motion during the cooling of the primitive planets, but there is a grave objection to supposing that they can have had any important effect on the orbits of the planets. Their orbits, like those of the planets, must have been initially highly eccentric. The gravitation of the planets would make their apsidal lines rotate in at most some thousands of years, and thus, even though they might be moving without collisions initially, they would, in a short time cosmogonically, reach a state where any region large enough to contain a moderate

number of them would contain nearly equal numbers moving inwards and outwards with velocities comparable with the velocity of a planet moving in a circular orbit in the same neighbourhood. Now if two solid bodies moving with such velocities collided, they would certainly be volatilized. Meteors, for instance, are volatilized when they enter the earth's atmosphere with such velocities, even in spite of the opportunity for loss of heat by convection and radiation during the several seconds the flight lasts. In the absence of air, the impacts of the bare surfaces would ensure that the whole of the energy lost on account of the imperfection of restitution would be liberated instantly, and thus, still more than in the case of meteors in the atmosphere, volatilization would ensue.

If now we suppose the planetesimals to be spheres, of radius c, the probability of any particular planetesimal striking another body, of radius a, is, for the same relative velocity, proportional to the square of the sum of the radii. Thus the probability of a planetesimal hitting any other particular planetesimal is to the probability of its hitting a planet of radius a in the ratio $4c^2/(a + c)^2$ or practically $4c^2/a^2$; and if we add up the probabilities for all planetesimals, the probability of a particular planetesimal hitting any other planetesimal in general is to that of its hitting a planet in the ratio of four times the surface of all the planetesimals to the surface of the planet. It can easily be seen that this result is still correct as regards order of magnitude when the planetesimals are no longer supposed spherical. The condition that more planetesimals were swept up by the planets than were volatilized is therefore that the total surface of the planetesimals was much less than that of the planets.

Now let us consider the effect of accretion on a planet moving in an eccentric orbit. When such a planet is near aphelion its velocity is less than that corresponding to a circular orbit at the same distance, while when it is near perihelion its velocity is greater than that for a circular orbit. If we suppose for the moment that the planetesimals were moving in circular orbits, the planet would therefore overtake them when near perihelion, and be overtaken by them when near aphelion. Now we know from the theory of impact that whenever two bodies unite into one the velocity of the combined body is between those of the original ones. Thus the difference between the velocities of the body and of the planetesimals in its neighbourhood would be diminished by every impact, and therefore its orbit would gradually become more nearly circular, just as in the case of a gaseous resisting medium. If we allow for the fact that the planetesimals were moving in highly eccentric orbits like those of the planets, the effect would be more marked, for a planet would pick up more particles it met than it overtook. We can therefore agree that the eccentricity would decrease. In this way the authors of the Planetesimal Hypothesis say that the eccentricities would be reduced to a fraction of their original values, and that the nearly circular orbits of the planets are explained. The explanation is, however, insufficient. If a planet picked up a planetesimal of its own mass, moving in a circular orbit, the difference between its velocity and that of a body describing a circular orbit would only be halved, and it is not difficult to show that the finely divided condition of the matter picked up does not alter this conclusion. Hence to produce a considerable reduction in the eccentricities of the planetary orbits the total mass of the planetesimals picked up must have been comparable with the masses of the planets themselves. But the more finely divided the

matter the greater would its surface be, and therefore the total surface of the planetesimals must have exceeded many times that of the planets. This result, by the last paragraph, shows that collisions between planetesimals must have been enormously more frequent than those between planets, and therefore the planetesimals must have been volatilized by collisions among themselves before they had time to affect the eccentricities of the planetary orbits appreciably. Thus the reduction of the eccentricities of the orbits of the planets by accretion is impossible.

A·12. This discussion has constructive value, in that it shows that the masses of the planets cannot have increased by any considerable fractions of themselves since they were formed. Thus all the meteors picked up cannot have had much effect on the earth's size or on its orbit. The solid particles that left the planets in their early history may have collided and volatilized each other; in that case they would have been added to the gaseous resisting medium, and may in that form have had some cosmogonical importance.

A·2. *Possible Compression.* Though the impact of solid planetesimals on the surfaces of the planets cannot have had much effect on the masses and motions of the planets, it is as well to examine other arguments that have been advanced against the theory of the former fluidity of the earth and in favour of the view of Chamberlin that the earth has always been solid. In the first place, Chamberlin has accepted the opinion of Osmond Fisher that thermal contraction is insufficient to account for mountain-building, and in place of it has suggested that accretion would give the requisite compression. Matter deposited on the surface of the earth would compress the interior, and in consequence of this contraction the interior would be too small for the exterior and would produce crumpling. In this way a linear crumpling comparable to the radius of the earth itself could, according to Chamberlin, be produced. It will be seen, however, that this could be true only if the earth's radius had been increased by accretion by a large fraction of itself, which has just been seen to be impossible. Further, the compression would be a slow process. The matter deposited on the surface would necessarily be in a fragmentary condition. It could only begin to be consolidated by pressure when it had been buried to such a depth that the pressure was sufficient to cause plastic deformation. Then gradual flow would commence, accompanied by reduction of volume. Meanwhile the matter above it, being under less pressure, has not begun to be consolidated. Contraction below therefore produces no crumpling; the fragments merely roll and slide over each other. Thus on the planetesimal theory there could at no stage be crumpling at the surface till accretion had ceased. Afterwards crumpling could arise only through internal cooling, which would necessarily be less than was possible in a fluid earth, and through the continuance of the consolidation of the interior. Now plastic flow once started in any region would continue until adjustment was complete, and would for ordinary substances be rapid in comparison with the slow process of accretion. Thus at most a few kilometres in depth would be capable of flow under the actual pressures existing, without having already adjusted themselves completely before the end of accretion. Consolidation could therefore contribute at most only a few kilometres to the available compression at the surface. The Planetesimal theory

is therefore able to explain less mountain-building than the fluid earth theory.

A·3. *The History of the Atmosphere.* It has also been argued that a fluid earth could not have retained any atmosphere, and especially water vapour. This argument is capable of two independent refutations: first, that the earth could have retained its water vapour, and second, that even if it had lost its whole atmosphere it could have produced a new one. It is seen that Chamberlin's argument supposes that if the earth's surface reached a temperature of 1500° A. most of the gases of the atmosphere would depart from the influence of its gravitation. If C is the mean square velocity in a gas, Jeans has shown that half the gas would be lost in 2×10^6 years if C was as high as 2·5 km./sec. For hydrogen (H_2) at 280° A., C is 1·9 km./sec. Thus hydrogen could certainly be retained at ordinary temperatures for a longer time than has been considered could have elapsed between the formation of the earth and the cooling of its surface. Now C is proportional to $m^{-\frac{1}{2}}V^{\frac{1}{2}}$, where m is the molecular weight and V the absolute temperature. For water vapour at 1500°, C would therefore be only 1·47 km./sec.; thus water vapour could be easily retained for the time required. Still more so oxygen and nitrogen would be retained. If the earth was primitively distended, water vapour could still be retained when the radius was three times as great as at present, since the critical velocity is proportional to the reciprocal of the square root of the radius. There is therefore no insuperable obstacle to the retention of water vapour and all heavier constituents of the atmosphere when the earth was fluid.

A·31. It is possible, however, that very little of the atmosphere is primitive. We know that hydrogen, water vapour, carbon dioxide, and nitrogen are being continually evolved from volcanoes, and it is possible that most of the atmosphere has been produced from this source since solidification. This requires all our water to have been within the earth initially. A few years ago the idea of water at 1500° mixed with siliceous constituents would have seemed ridiculous, but it is now known that at high temperatures and pressures water and rocks actually mix freely in all proportions. Thus it is probable that all the water in the primitive atmosphere was absorbed by the rocks and has only been given off again since solidification*. It was for a long time considered difficult to explain how igneous rocks came to contain water of crystallization, but this problem may now be regarded as solved.

Oxygen is not produced to any appreciable extent by volcanoes, but carbon dioxide is, and it is possible that all the oxygen in the atmosphere has been formed from carbon dioxide by the action of plants. There is a difficulty in this suggestion, though perhaps not an insuperable one. The capacity to assimilate carbon is confined to the plants that contain chlorophyll, which are in general the higher plants. The lower plants absorb oxygen and expire carbon dioxide like animals; so do even the green plants when in the dark. Thus it is not easy to see how the conversion of carbon dioxide into oxygen could have commenced until plant life had reached a fairly advanced stage in evolution, and until that stage was reached the plants must have lacked the oxygen they needed. If this objection is valid the oxygen in the atmosphere, or at any rate some of it, must be primitive.

* J. W. Evans, *Observatory*, **42**, 1919, 165–7.

Hydrogen is produced in quite sufficient quantities to account for the trace of it present in the atmosphere; most of it combines soon with oxygen. Helium is produced by radioactivity. The difficulty in this case is to understand why so little of it is present in the atmosphere. With the data of Chapter VI we can show that average acid rock produces helium at the rate of $1{\cdot}7 \times 10^{-12}$ c.c. of helium per gram of rock per year. The helium at the earth's surface is responsible for about a millionth of the atmospheric pressure; thus if all the helium in the atmosphere were under ordinary atmospheric pressure it would make a layer over the surface 0·8 cm. thick. If now 1000 c.c. of acid rock were denuded from each square centimetre of the earth's surface, and all their helium transferred to the atmosphere, they would give 5×10^{-9} c.c. helium per cm.^{2} of the surface for each year of the interval between the formation of the rock and its denudation. If then the average age of igneous rocks when denuded was 160 million years, all the helium of the atmosphere would be explained if the average thickness of igneous rocks denuded from the earth's surface was 10 metres. This is an impossibly small amount. It could be increased if a large fraction of the helium in rocks is retained in the sedimentary rocks formed from them, but it appears unlikely that this fraction exceeds $\frac{1}{2}$, and thus the depth can only be doubled. No satisfactory explanation of why the atmosphere contains so little helium has yet been offered.

On the theory of the former fluidity of the earth it is therefore possible to account for the gases of the atmosphere. On Chamberlin's theory, however, it is not possible to account for the existence of the atmosphere. The small gravitative power of his small earth nucleus would have made it unable to retain permanent gases during its aggregation, and any gases or water absorbed in it could never have got out again, since it was by hypothesis buried below some thousands of kilometres of planetesimals. The atmosphere must then have been brought by the planetesimals. But the planetesimals, like modern meteorites, must have been completely arid and atmosphereless. Thus, as in the previous case, Chamberlin's arguments are more injurious to his own theory than to the one he is attacking.

A·4. *The Primitive Crust*. It has been suggested that the primitive crust of a formerly solid earth should be in a characteristic condition, and might be geologically recognizable. No trace of the primitive crust has been found, and some geologists have made this fact the basis of an attack on the theory of the former fluidity of the earth. In view, however, of the extent of denudation and redeposition that we know to have occurred, there would be no cause for surprise if the whole of the primitive crust should have been removed or buried. The argument, so far as it has any validity, is, of course, equally applicable against the planetesimal theory, for planetesimal matter unaltered since deposition would also be in an easily recognizable condition, and has not been identified.

A·5. *The Land and Water Hemispheres*. A further argument has been based on the existence of the land and water hemispheres. One of the outstanding problems of geophysics is the mechanism that produced the marked asymmetry of the distribution of land and sea. No asymmetry about the axis of rotation could exist in a fluid earth, though it would be premature to say that no such asymmetry could be developed under the

stresses acquired at solidification. It has been suggested, however, that the phenomenon is explicable on the planetesimal theory. More planetesimals might have fallen on one side of the earth than on the other, and thus have produced the observed inequality. The probability of such a hypothesis is, however, inappreciable.

Let the radius of the earth be a, and consider the probability of any planetesimal falling in a range of longitude $2\pi\alpha$ of the surface. This probability is evidently α; for on account of the rotation of the earth, all meridians are equally frequently presented to the planetesimals, in whatever direction they may approach. Suppose then that n planetesimals fall altogether. The prior probability that m of them fall within the sector $2\pi\alpha$, the fall of each being unaffected by those that have fallen already, is

$$^nC_m\,(1-\alpha)^{n-m}\,(\alpha)^m \qquad \text{.........(1)},$$

where ^nC_m denotes the number of combinations of n things taken m at a time. Let us denote this by P. This is a maximum if $m/n = \alpha$. Thus a uniform distribution in longitude is the most probable. Let us now investigate the probability of a departure from the uniform distribution by a given amount. Put

$$m/n = \alpha + \xi \qquad \text{.........(2)},$$

where ξ is small and $n\xi^2$ moderate. Then we may approximate by Stirling's formula

$$\log n! = -n + (n + \tfrac{1}{2}) \log n + \tfrac{1}{2} \log 2\pi \qquad \text{.........(3)},$$

and obtain, when n is large and neither α nor $1-\alpha$ very small,

$$\log P = -\tfrac{1}{2} \log 2\pi n\alpha\,(1-\alpha) - \frac{n\xi^2}{2\alpha\,(1-\alpha)} \qquad \text{.........(4)}$$

$$+ \text{ terms of order } \xi \text{ and } \xi^2.$$

Now if we consider the range between $n\,(\alpha+\xi)$ and $n\,(\alpha+\xi+d\xi)$, where $d\xi$ is small compared with $n^{-\frac{1}{2}}$, the number of possible values of m in this range is clearly $n\,d\xi$. Thus the probability that m lies in this range is

$$\left\{\frac{n}{2\pi\alpha\,(1-\alpha)}\right\}^{\frac{1}{2}} e^{-n\xi^2/2\alpha(1-\alpha)}\,d\xi \qquad \text{.........(5)}.$$

The probability that m exceeds $n\,(\alpha+\xi)$ is therefore

$$Q = \left\{\frac{n}{2\pi\alpha\,(1-\alpha)}\right\}^{\frac{1}{2}} \int_\xi^\infty e^{-n\xi^2/2\alpha(1-\alpha)}\,d\xi$$

$$= \tfrac{1}{2}\left\{1 - \operatorname{Erf} \xi \left(\frac{n}{2\alpha\,(1-\alpha)}\right)^{\frac{1}{2}}\right\} \qquad \text{.........(6)}.$$

Now when the argument is great, an approximation to $\operatorname{Erf} x$ is given by the formula

$$1 - \operatorname{Erf} x = \frac{e^{-x^2}}{x\sqrt{\pi}} \qquad \text{.........(7)},$$

so that

$$Q = \left(\frac{\alpha\,(1-\alpha)}{2n\pi}\right)^{\frac{1}{2}} \frac{1}{\xi} e^{-n\xi^2/2\alpha(1-\alpha)} \qquad \text{.........(8)}.$$

In our problem n is the total number of planetesimals picked up by the earth. If then b be the radius of a planetesimal, and if the radius of the earth have increased by accretion by $\frac{1}{4}$ of its original value, we have

$$n = a^3\{1 - (\tfrac{4}{5})^3\}/b^3 = 0{\cdot}5a^3/b^3 \text{ nearly} \qquad \text{.........(9)}.$$

The Pacific Ocean is, on an average, about a kilometre deep, and covers about half the earth. Thus if we consider the probability of getting such a depth by accident, as has been suggested, we take $\alpha = \frac{1}{2}$ and $\xi = \frac{1}{1000}$. Thus

$$Q = \left(\frac{b^3}{4\pi a^3}\right)^{\frac{1}{2}} 1000\, e^{-a^3/10^6 b^3} \qquad \text{.........(10).}$$

Evidently the smaller the planetesimals the smaller is the probability of the hypothesis. We can hardly suppose the planetesimal to have a larger radius than a kilometre, in which case $a/b = 6000$. Then

$$Q = \frac{1}{1500} e^{-2\times 10^5} \qquad \text{.........(11),}$$

which is utterly insignificant. Thus the formation of the land and water hemispheres in the process of accretion is practically impossible.

Chamberlin suggests in *The Origin of the Earth* that denudation during growth would increase the asymmetry, even though this might be originally insignificant. This does not, however, appear likely. If an ocean formed early on a nearly spherical nucleus, as he suggests, it would submerge any small inequality that might arise, and the only denudation possible would be the insignificant amount that goes on at the ocean bottom. This hypothesis does not therefore appear to solve any difficulty presented by the fluid earth theory.

A·6. I believe, therefore, that the Planetesimal Hypothesis does not offer a solution of any of the great outstanding problems of geophysics, and that on cosmogonical grounds it is quite unacceptable. It has, however, been of considerable value in the past in suggesting new lines of attack on our problems, and the tidal theory here adopted arose as a modification of the Planetesimal Hypothesis designed to avoid the objections that appeared fatal to it in its original form. To reject this hypothesis now is not to deny its importance in the history of cosmogony and geophysics.

APPENDIX B

Jeans's Theory

"It is an old maxim of mine that when you have excluded the impossible, whatever remains, however improbable, must be the truth."

A. CONAN DOYLE, *The Adventures of Sherlock Holmes.*

B·1. The chief point of difference between the theory of Chapters II, III and IV and that of Jeans is that whereas I suppose the primitive sun at the time of the disruption to have been well within the orbit of Mercury, he supposes it to have been distended so as to include the orbit of Neptune. There are three serious objections to such a distension of the sun. First, it could, according to modern theories of the constitution of stars, have had an effective temperature of only 200° absolute, and therefore if it could have existed at all it would have consisted of solid dust. Thus the theory of the evolution of the filament on the hypothesis that it was initially gaseous would break down completely. Secondly, it implies an extremely low density, and hence by the formula 2·4 (9) the velocity of sound in a gaseous filament would have to be rather low to account for masses as small as those of the planets. Thirdly, planets produced at distances from the sun greater than the present distance of Neptune would have had to be brought in by some means. The only agent available for this purpose appears to be a resisting medium. It has been shown in 4·2 that a resisting medium, the motion at any point of which was not sensibly different from that of a planet in a circular orbit at the same point, could have no effect on the mean distance of a planet; and that a medium for which this difference was considerable would have too small a density at such a distance to produce any appreciable frictional effect on a planet. Thus in neither case could the mean distances be affected appreciably by a resisting medium.

B·2. *Prior Probability of a Disruptive Encounter.* Jeans's reason for adopting such a distension seems to be based on the belief that the prior probability of any encounter close enough to break up the sun when less distended is so small as to forbid it. This, however, does not seem to be the case. Jeans estimates (*loc. cit.* p. 279) that in the present universe the average interval between encounters at a distance less than $1{\cdot}3 \times 10^{15}$ cm. is about 3×10^{10} years. Thus the probability that in any one year any particular star that we have no other data about will have an encounter at this distance or less is $1/(3 \times 10^{10})$. This probability is proportional to the square of the distance of approach; therefore the probability that any one star would have one at a distance of 3×10^{12} cm. is $1/(6 \times 10^{15})$. Hence the probability that any particular star would have an encounter within the latter distance in 2×10^7 years, a time probably comparable with the time the sun spent in the giant stage, is $1 - \{1 - 1/(6 \times 10^{15})\}^{2\times10^7} = 1/(3 \times 10^8)$, and the probability that one star in the whole universe, containing perhaps 10^9 stars, would have an encounter at this distance, is $1 - \{1 - 1/(3 \times 10^8)\}^{10^9}$ or practically unity. Thus the theory makes it practically certain that at least one star in the universe has had an encounter at the distance here assumed, and our only datum in the matter is that there is at least one solar

system. Thus the hypothesis that the diameter of the sun at the encounter was only a fraction of that of the orbit of Mercury is consistent with knowledge of stellar motions. It is, however, probable that systems of the type of the solar system are the exception in the universe and not the rule.

B·21. It may be remarked that the tidal theory of the origin of the solar system might have a considerable probability on the evidence before us, even if stellar encounters were much less frequent than has been inferred from the present condition of the stellar system. For, let, in the notation used by Dr Wrinch and myself*, h denote the aggregate of propositions believed independently of experience, and let $P\ (p : qh)$ denote the probability that a proposition p is true, given that h and another proposition or set of propositions q are true. Let $\sim p$ denote the proposition that p is untrue, and pq the proposition that p and q are both true. Let the empirical data of physics, not including the existence of the solar system, be denoted by k, and let the proposition that the solar system exists be denoted by m. The laws of physics, which we may denote by l, have a high probability on the data k, so that

$$P\ (l : kh) = 1 - \alpha \qquad \text{.........(1)},$$

where α is very small.

The arguments for the tidal theory fall into two groups. First, it is shown that, given our physical laws, no theory other than a tidal one can account for the solar system as it is. If then t denotes the proposition that the initial conditions required by the tidal theory once existed, this result may be expressed by

$$P\ (m : \sim t\,.\,lkh) = \beta \qquad \text{.........(2)},$$

where β is very small. It would be zero if the proof possessed strict mathematical accuracy, and the proof available is enough to show that β must be exceedingly small.

Next, it is shown that the tidal theory is capable of explaining the main features of our system, given suitable initial conditions. Thus

$$P\ (m : tlkh) = \gamma \qquad \text{.........(3)},$$

where γ is certainly not small and may be nearly unity.

Now we have the general propositions of the theory of probability

$$P\ (pq : h) + P\ (p\,.\sim q : h) = P\ (p : h) \qquad \text{.........(4)},$$

$$P\ (pq : h) = P\ (p : h)\, P\ (q : ph) \qquad \text{.........(5)}.$$

Then

$$\begin{aligned} P\ (m : kh) &= P\ (mlt : kh) + P\ (ml\,.\sim t : kh) + P\ (m\,.\sim l : kh) \\ &= P\ (l : kh)\, P\ (t : lkh)\, P\ (m : tlkh) \\ &\quad + P\ (l : kh)\, P\ (\sim t : lkh)\, P\ (m : \sim t\,.\,lkh) \\ &\quad + P\ (\sim l : kh)\, P\ (m : \sim l\,.\,kh) \qquad \text{.........(6)}. \end{aligned}$$

Let us write

$$P\ (t : lkh) = \tau,$$

$$P\ (m : \sim l\,.\,kh) = \delta.$$

Then (6) becomes

$$P\ (m : kh) = (1 - \alpha)\,\{\tau\gamma + (1 - \tau)\,\beta\} + \alpha\delta \qquad \text{.........(7)}.$$

* *Phil. Mag.* **38**, 1919, 717–731.

Again, $$P\,(mlt : kh) = (1 - \alpha)\,\tau\gamma,$$
being the first term in (6), and also
$$= P\,(m : kh)\,P\,(lt : mkh) \qquad \text{.........(8).}$$
Combining (7) and (8), we have
$$P\,(lt : mkh) = \frac{(1 - \alpha)\,\tau\gamma}{(1 - \alpha)\,\{\tau\gamma + (1 - \tau)\,\beta\} + \alpha\delta} \qquad \text{.........(9).}$$
Now δ, being a probability number, is necessarily not greater than **1**, and α and β are exceedingly small. Thus the expression on the right will be practically unity unless $\tau\gamma$ is so small as to be comparable with either α or β. Omitting this alternative for a moment, we see that $P\,(lt : mkh)$ is practically unity. In other words, given the empirical data of physics and the existence of the solar system as it is, it is practically certain both that the laws of physics used in the argument are true and that the initial conditions required by the tidal theory once occurred; and therefore that the tidal theory is true.

This argument would break down if $\tau\gamma$ was very small. We have seen in B·2 that τ is probably nearly unity, and that γ is moderate, and therefore there is no question of its failure with the data here used. For failure it would be necessary for $\tau\gamma$ to be less than one of the two quantities α and β. In other words, the probability of there having ever been, during the giant stage of a single star in the universe, an encounter at a distance less than 3×10^{12} cm., must be less than either the probability of a fallacy in the argument against the possibility of the formation of the solar system by purely internal action, or the probability that the laws of physics are wrong This gives a conception of the strength of the position of the tidal theory. If β is the greatest of the quantities $\gamma\tau$, α and β, $P\,(m : kh)$ will be practically
$$(1 - \alpha)\,(1 - \tau)\,\beta = P\,(ml\,.\sim t : kh)$$
$$= P\,(m : kh)\,P\,(l\,.\sim t : mkh),$$
and therefore $P\,(l\,.\sim t : mkh)$ will be practically unity. Thus it will be practically certain that the solar system was formed by some non-tidal action and that the laws of motion are right. If, again, α is the greatest, it will follow that the laws of motion are probably wrong. Thus we see that the extreme smallness of $\gamma\tau$, besides forcing us to reject the tidal theory, would make other much more bizarre inferences necessary.

The tidal theory therefore appears to be in a stronger position than Jeans has shown.

APPENDIX C

The Hypothesis of the Indefinite Deformability of the Earth by Small Stresses

"If ye have faith as a grain of mustard seed, ye shall say unto this mountain, Remove hence to yonder place; and it shall remove; and nothing shall be impossible unto you.

"Howbeit this kind goeth not out but by prayer and fasting."

St Matthew xvii. 20–21.

C·1. In many discussions by geologists of the physics of geological processes, it is freely assumed that any stress, however small, can deform the earth to any assignable extent, provided only that it acts long enough. In the language of Chapter IX, the earth is a plastic body of zero strength, or is liquevitreous; many writers would call such a substance a fluid, but for reasons already given I prefer to class it with solids, while noticing the special qualities that are associated with the absence of strength.

It has been seen from thermal considerations that this description is probably applicable to the matter of the earth at depths greater than 700 km. From this result the perfection of the compensation of the continents and of the second harmonic inequality in the figure of the earth have been inferred, and the latter of these inferences has been accurately verified by geodesy. This weakness also plays an important part in the theory of the compensation of mountain ranges.

The authors considered, however, suppose the complete absence of strength to extend up into the asthenosphere and even higher. In these regions, according to the theory already developed, the rocks have cooled some hundreds of degrees since solidification, and it would be natural to suppose, in the absence of more direct evidence, that they have a finite strength. The geodetic evidence, again, has been seen to show that the horizontal extent of the regions where the Hayford anomaly is systematically in one direction is some 700 km., suggesting a finite strength at depths of 120 km. and over. In the layer of compensation, again, the strength must be great, for otherwise the mountains would collapse under their own weight; it has been shown in 8·6 that greater strength in the upper layers is required to support a compensated mountain than an uncompensated one.

C·11. The question at issue is not, then, whether the whole earth has a finite strength or a zero strength; we know that neither of these alternatives is correct. It is purely a question of the quantitative determination of the depth where strength first becomes inappreciable. To evaluate this more definitely than has been done in 9·64 does not appear to be possible with the data at our disposal; which is only another way of saying that we know from geodesy nothing contradictory to the hypothesis that the strength is finite, though small, to a depth of many hundreds of kilometres.

C·12. The finite strength of the upper layers offers a formidable obstacle to any permanent deformation of the outer layers where the stresses involved are not very large. We know that the crust in the continents is

strong enough to hold up Mount Everest, and in the ocean bottom can hold down the Tuscarora Deep, and that the strength required to hold up an inequality depends almost wholly on its vertical and not on its horizontal extent. Thus the whole weight of the continents will not be enough to produce permanent deformation in the upper layers; much less will the small fraction of it that acts tangentially, on account of the asymmetry of the earth's figure. We cannot therefore accept hypotheses of the widespread migration of continents, unless forces enormously greater than any yet suggested are shown to be available.

C·2. A further impossible hypothesis has often been associated with hypotheses of continental drift and with other geological hypotheses based on the conception of the earth as devoid of strength. This is, that the small force can not only produce indefinitely great movement, given a long enough time, but that it can overcome a force many times greater acting in the opposite direction for the same time. In Wegener's theory, for instance, not only is a tiny force supposed to have moved America right across the present Atlantic, but the resistance of the Pacific floor to the motion is supposed to have uplifted the Rocky mountains. Now, given a sufficiently weak earth and enough time it might be possible to twist the outside of the earth over the inside to any extent. So long as the layers of equal density remained symmetrical about the polar axis, no elevation or depression of rocks taking place, deformation could proceed undisturbed, America going steadily on its way without mountain-building or any other phenomenon observable by geologists. In order that mountain-building may take place, however, energy must be supplied to raise and lower the rocks affected against gravity; and the stress available must overcome gravity, and therefore must exceed the pressure due to the weight of the mountain. Tidal friction and differences between the values of gravity at the tops and bottoms of continents are the agencies usually considered in theories of this type; they are capable of producing stresses of the order of 10^{-5} dynes per sq. cm., whereas to elevate the Rockies about 10^9 dynes per sq. cm. would be required.

The assumption that the earth can be deformed indefinitely by small forces, provided only that they act long enough, is therefore a very dangerous one, and liable to lead to serious errors.

APPENDIX D

Theories of Climatic Variation

"The winter of our discontent."
SHAKESPEARE, *King Richard III*, Act I, Sc. 1.

D·1. *The Geological Evidence for Climatic Changes.* Climatic variation as such is essentially a meteorological topic, and therefore is outside the scope of the present work. On the other hand, climatic variations have produced very great effects on the surface features of the earth in the past, and their causes may be largely drawn from the non-meteorological parts of geophysics. Since they serve as a connecting link between two problems definitely within our scope, some discussion of them may be in place.

The evidence indicating that considerable changes in climate have taken place in past geological ages is very considerable. The last glacial period, during which most of Northern Europe and America were buried under a thick ice sheet, is the best known of these vicissitudes; but it is not so well known that a glacial period with many similar features occurred in the Pre-Cambrian era, before the oldest known fossiliferous rocks were laid down, and another in the Permian period, about 300 million years ago, besides occasional local glaciations at other times. Between these glaciations there were mild intervals, such as the long spell of warmth in the Secondary and early Tertiary. At present the climates over most of the earth appear to be becoming warmer and drier, though there are places where this does not hold.

D·2. *Suggested Explanations.* Numerous attempts have been made to explain such facts. Most of the hypotheses offered may be classified under the following heads:

A. Changes in the motions of the earth as a whole, especially in
   (1) the eccentricity of its orbit;
   (2) the inclination of its axis of rotation to the ecliptic.

B. Changes in the composition of the atmosphere and ocean, especially
   (1) the amount of carbon dioxide;
   (2) the amount of volcanic dust;
   (3) the amount of salt in the sea.

C. Changes in topography:
   (1) in the area of the continents;
   (2) in the height of the continents;
   (3) in the distribution of the continents with regard to one another and to the polar axis.

D. Changes in the sun's radiation.

E. Changes in the internal heat of the earth.

F. The passage of the solar system through cold regions of space.

D·3. Of these, A, E, and F may be rejected without difficulty. First, changes in the eccentricity formed the basis of Croll's famous theory, which required that glaciation should occur when the eccentricity was a maximum, and should alternate between the northern and southern hemispheres in the period of the precession of the equinoxes, the glaciated hemisphere being the one containing the pole presented to the sun at perihelion. Both the variation of the eccentricity and the precession of the equinoxes are purely dynamical in origin and extremely regular. Thus glacial periods should have alternated at intervals of less than a million years throughout geological time, which is very far from the state of affairs outlined in the second paragraph above. Changes in the inclination of the axis may arise in two ways, first, by planetary perturbations of the plane of the earth's orbit, and second, through internal deformations of the earth itself. The first type of change is regular, and therefore fails for the same reason as Croll's hypothesis; while Sir G. H. Darwin showed that changes in the inclination through internal deformation can never be considerable. The inclination may have undergone some change through tidal friction, but such change would always be in the same direction, whereas the climatic changes have been oscillatory.

D·31. Hypothesis E fails, because, as shown in Chapter VI, the surface temperature of the earth must have been almost wholly maintained by solar radiation practically ever since it became solid at the surface, and certainly throughout geological time. Conduction from the interior is in comparison quite unimportant.

D·32. Strictly speaking, Hypothesis F is meaningless; there is no such thing as the temperature of space, for temperature is essentially a property of matter. It might, however, be interpreted to mean the temperature that a given body would take up if it was exposed to the radiation in the neighbourhood considered; so that it would be the temperature such that the given body radiates as much heat as it receives. But at present the radiation received from other sources is a very small fraction of that received from the sun, and therefore no possible reduction in it could make any appreciable difference to terrestrial temperatures.

D·4. The phenomena enumerated under B must all have occurred to some extent, but it is unlikely that any of them is among the main causes of climatic variation. It has been suggested by W. J. Humphreys* that great volcanic eruptions, such as those of Krakatoa (1883) and Katmai (1912), may produce a lowering of temperature over the earth as a whole for some months; thus volcanic activity on a still larger scale might produce a glacial period. It does not appear, however, that volcanic activity and glaciation have been closely correlated in past epochs. If they had been, great glaciations would have occurred during the times of greatest activity in the Ordovician, Devonian, and Tertiary periods, which is not the case. The absorption of dark heat radiation by carbon dioxide may be appreciable, but it appears probable that the conditions have always been such that any radiation absorbed by CO_2 would have been equally effectively absorbed by water vapour in the absence of the former, so that the thermal effect of variations in the amount of CO_2 in the atmosphere can hardly be great. Variations in the salt content of the ocean have

* *Journ. Franklin Inst.* **176**, 1913, 131.

been considered important by Chamberlin. Variation of the density of the ocean from place to place must have some effect in producing a general circulation of the ocean. Thermal expansion near the equator reduces the density of the water, and must therefore have such an effect. If, however, the ocean were much more salt, evaporation of water near the equator would raise the density more than thermal expansion reduces it, and the density circulation would be reversed. This is considered capable of producing an important effect on climate, since the atmosphere is largely heated by contact with the sea. The difficulty in this theory is that the surface waters do not move with the deeper parts of the ocean, but are driven along by the wind*; with the same wind, their motion is practically independent of the density distribution. It is the surface waters alone that affect the temperature of the air. Thus the system of the surface waters and the air is practically self-contained, and the temperature of the air cannot be affected by the salinity of the sea.

D·5. Phenomena of group C must contribute considerably to the variation of climate. It is known from meteorology that changes in topography could have an important effect on climate, and from geology that such changes have actually taken place; what is uncertain is, first, how far known changes in topography can account for the changes of climate indicated by the geological record, and, second, in cases where prior knowledge of the putative causes is so small that we have to make *ad hoc* hypotheses concerning their time and extent, how numerous are the observed climatic variations that agree with those inferred from the hypotheses.

D·51. The chief work on these lines has been done by C. E. P. Brooks†. His method is based on the correlations between the normal temperatures for January, July, and the whole year, on the one hand, and the nature of the surroundings on the other, found for a network of places over the whole earth. The places he considers are regularly distributed at equal intervals of latitude and longitude all over the earth. Around each he draws a circle of radius 10° of latitude, or 600 nautical miles. The percentages of land and ice in the semicircles to the west and east of the station are recorded separately, and the correlations between these and the temperatures are calculated. From these correlations it is possible to make a quantitative estimate of the effect on temperature of continentality and the direction of the prevailing winds. The method is somewhat crude, but certainly takes into account the main features of the problem. The main results are what might be expected from our general knowledge, that land in the neighbourhood of a station makes for extremes of temperature, that ice produces lower temperatures than unglaciated land, and that land or ice produces more effect when it is to windward than when it is to leeward. The special feature of Brooks's work is that it is quantitative as well as qualitative, and can therefore be applied to find the distributions of temperature for other distributions of land and sea. The temperature distributions for other geological dates than the present can therefore be found. Brooks uses temperatures reduced to sea level, so that the effect of altitude must be added before the actual temperature at a station is determined.

* Jeffreys, *Phil. Mag.* **39**, 1920, 578–586.

† *The Evolution of Climate*, Benn Bros. 1922.

In this way Brooks is able to show that more oceanic conditions, which actually existed, are quantitatively able to account for the mild climate of the Eocene period. A general elevation of the land proceeded throughout the Tertiary Era, and when the Scandinavian highlands and the Rocky Mountains reached the snow line an ice sheet commenced to form. Ice radiates more and absorbs less heat than unglaciated land at the same temperature, and it is for this reason that it becomes colder than unglaciated land exposed under similar conditions, and produces a greater cooling effect in its neighbourhood. Wind blowing over an unglaciated mountain range reaches the same level on the other side without having been cooled much, or may even have been somewhat heated by the Föhn effect; but wind blowing over an ice sheet is very much cooled when it gets to its original level again. Thus whereas the rising mountains did not produce much disturbance of temperature to leeward before they reached the snow-line, they would afterwards produce a substantial depression. Precipitation upon them would fall as snow instead of rain. That falling on the windward side would form glaciers, and return quickly to the sea; but that to leeward would have no easy outlet, and would accumulate. Thus the ice sheet would tend to spread, especially to leeward. The increased precipitation consequent on the reduction of the temperature of the land would play an important part in the thickening of the ice sheet. The sheet would not necessarily have its highest point above the original mountains; the precipitation to leeward might well raise the surface of the ice to above the original mountain tops, and then the highest point would proceed to move steadily to leeward. The actual events during the glacial period and afterwards agree closely with Brooks's inferences. In particular, some sand dunes in North Germany, formed at this time, have their tips pointing to the west instead of the east, showing that the prevailing wind there at the time was from the east. This is exactly what would be expected from the presence of the Scandinavian ice sheet, which would produce a permanent area of high pressure in Scandinavia, and therefore east winds over Germany. In many other parts of the world striking agreements are found.

Brooks's theory is therefore a very substantial contribution to our understanding of climatic change; but it does not furnish a complete explanation. Glaciation did not begin in many places until long after the mountain tops were not only above the snow line, but above the height where there was any considerable precipitation at all; this is shown by the presence of unglaciated mountain tops, called 'nunataks,' in Scandinavia and the Rockies. It appears as if the later stages, at least, of the elevation of the mountains took place under conditions when the snowfall was inappreciable, and that the ice sheet did not begin to form until some further change of climate, not attributable to the mountains, had supervened. The Cambrian, Ordovician, and Silurian folds, again, must have raised mountains quite comparable with those of the Tertiary, but do not appear to have been followed by glaciation on anything like the same scale, again suggesting that mountain formation, though it may be a necessary preliminary to glaciation, is not a sufficient condition for it. The crustal movements in the last thousand years, again, do not appear enough to account for the climatic changes during that interval.

D·6. Changes in the situation of the polar axis with regard to the land have often been summoned to explain climatic changes that we have no

satisfactory explanation for. In particular, Wegener has attributed the Permian glaciation in several parts of the Southern hemisphere to this cause. He supposes that at that time South America, South Africa, India, and Australia were all united, meeting somewhere in the South Indian Ocean, and that they have since drifted apart. The presence of a glacial flora in all these places at that time is then explained by the *ad hoc* hypothesis that the south pole was near the junction. The argument of Appendix C seems in itself fatal to this theory. Further, it has been pointed out by Lake* that a similar glaciation took place at this time in Northern Baluchistan, which would on Wegener's hypothesis have been practically on the equator.

D·61. A displacement of the exterior of the earth bodily over the interior is not intrinsically impossible, in view of the weakness of the interior. It would not alter the inclination of the axis of resultant angular momentum to the ecliptic, and this axis would always practically coincide with the polar axis. So long as the thickness of the outer shell was small compared with the radius, the axis of resultant angular momentum of the earth as a whole would be nearly that of the interior; thus the poles would still be over almost the same points of the interior, and therefore different points of the exterior. A displacement of this type would produce important climatic changes; but so far no agency capable of producing it has been suggested.

D·7. We now come to the effects of changes in the sun's radiation. We have no means, apart from terrestrial climate, of knowing how the sun's radiation may have changed during geological time, and accordingly any hypothesis of this type is very difficult to test, and therefore very difficult either to prove or to disprove. If any changes were due to this cause, the effects all over the earth should be such as can be attributed to the same change of solar radiation, and this might be thought to afford a test; but unfortunately we cannot determine the dates of events at widely separated places so accurately that we can be sure that the climatic variations at all of them actually took place at the same time; and further, it is far from certain that an increase in the intensity of solar radiation would affect temperatures all over the world in the same direction. W. Köppen† found that the temperature of the earth is highest about sunspot minimum, when the sun is radiating least. This paradoxical result received a partial explanation from H. F. Blanford‡. His suggestion is that high radiation increases the evaporation over the ocean, and therefore increases the cloudiness and rainfall over the land. On both the latter grounds the temperature of the land is reduced. It is to be observed that the temperature of the sea must be raised, for if it were lowered the evaporation would be less than normal, and the argument would have cut off its own feet. If the theory is correct, it can refer only to a limited range of conditions, for it is plain that if the sun's radiation were zero the temperature of the earth would be practically absolute zero, and if it were great enough the oceans would boil and the temperature of the land would be raised to boiling point.

* *Geog Journ.* 1923.

† *Zs. Osterr. Ges. Meteor.* 8, 1873, 241–252.

‡ *Nature*, 43, 1891, 583.

The correlations of terrestrial temperatures and rainfall with solar radiation are certainly small; a summary of the results obtained, with references to the literature, is given by C. E. P. Brooks in the *Meteorological Magazine* for June 1921. A connexion between sunspots and cyclones in the South Indian Ocean has been traced by Meldrum, but cyclones in other oceans show little connexion with sunspots or with one another*.

E. Huntington and S. S. Visher† have given an account of a theory of climatic changes, attributing most of the changes somewhat naïvely to variations in solar radiation. It certainly seems that some of the changes are most probably due to such variations, more on account of the failure of other hypotheses than on account of any conspicuous successes of this one. There is so far no satisfactory explanation of why the sun's radiation should have varied in the past; but as we do not know why it varies at present this offers no serious objection. Huntington and Visher offer an explanation based on the hypothesis that extra solar radiation is stimulated by variations in the radiation received by the sun from the stars; but this can be rejected, for if it were so, the components of a double star with a highly eccentric orbit should stimulate each other to such an extent as to make such stars strongly variable in a period equal to the period of revolution, which is not the case.

* Mrs G. F. Newnham, *Geophys. Mems. of the Meteorological Office*, No. 19, 1922.

† *Climatic Changes, their Nature and Causes.* Yale Univ. Press, 1922.

APPENDIX E

Empirical Periodicities

"Backward or forward, it's just as far;
Out or in, the way's as narrow." IBSEN, *Peer Gynt.*

E·1. Methods of harmonic analysis have been applied to many of the data of geophysics, especially in seismology, terrestrial magnetism, and meteorology. The principle of the method is as follows.

Suppose we have $2n - 1$ observations of a variable x, distributed at equal intervals of another variable t. As a rule t is the time. We may measure t from the first observation. Let the interval be T. Then the value of t at the $(p + 1)$th observation is pT, and at the last is $2(n - 1)T$. Let the value of x when $t = pT$ be x_p. Put

$$\theta = \frac{2\pi t}{(2n - 1)T} \qquad \text{.........(1).}$$

Consider the function

$$a_0 + a_1 \cos\theta + b_1 \sin\theta + \dots + a_r \cos r\theta + b_r \sin r\theta + \dots$$
$$+ a_{n-1}\cos(n - 1)\theta + b_{n-1}\sin(n - 1)\theta \qquad \text{.........(2),}$$

where the a's and b's are unknown constants, $2n - 1$ in number. If we suppose this function equal to x, our $2n - 1$ known values of x give $2n - 1$ linear equations to determine these coefficients, which can therefore be found uniquely provided the equations are determinate. We notice that as t increases by T, θ increases by $2\pi/(2n - 1) = \alpha$, say. The equation given by x_p is then

$$a_0 + a_1 \cos p\alpha + b_1 \sin p\alpha + \dots + a_r \cos rp\alpha + b_r \sin rp\alpha + \dots$$
$$+ a_{n-1}\cos(n - 1)p\alpha + b_{n-1}\sin(n - 1)p\alpha = x_p \text{...(3).}$$

There are equations of this type for all integral values of p from 0, 1 to $2n - 2$. Multiplying them in turn by

$$\cos 0,\ \cos r\alpha,\ \cos 2r\alpha,\ \dots \cos pr\alpha,\ \dots \cos(2n - 2)r\alpha,$$

and adding, we see that the coefficient of a_s in the sum is

$$\sum_{p=0}^{2n-2} \cos rp\alpha \cos sp\alpha = \tfrac{1}{2}\sum_{p=0}^{2n-2}\{\cos(r - s)p\alpha + \cos(r + s)p\alpha\} \text{......(4).}$$

Several cases arise. If r and s are unequal, each sum is of the form

$$\tfrac{1}{2}\sum_{p=0}^{2n-2}\cos mp\alpha = \frac{\sin\dfrac{2n-1}{2}m\alpha\cos(n - 1)m\alpha}{2\sin\frac{1}{2}m\alpha} \qquad \text{.........(5),}$$

and the first factor in the numerator is $\sin m\pi$, or zero. Thus in this case the coefficient is zero.

If r and s are equal but not zero, the second sum vanishes. The first is evidently $\frac{1}{2}(2n - 1)$.

If r and s are both zero, each sum is $\frac{1}{2}(2n-1)$ and therefore the two together are equal to $2n-1$. Hence

$$(2n-1)\,a_0 = \sum_{p=0}^{2n-2} x_p \qquad \text{.........(6)},$$

$$\tfrac{1}{2}(2n-1)\,a_r = \sum_{p=0}^{2n-2} x_p \cos rp\alpha \qquad \text{.........(7)}.$$

Similarly we can show that

$$\tfrac{1}{2}(2n-1)\,b_r = \sum_{p=0}^{2n-2} x_p \sin rp\alpha \qquad \text{.........(8)}.$$

Thus all the coefficients are determinate, and we have found the coefficients $a_0 \ldots\ldots a_{n-1}$, $b_1 \ldots\ldots b_{n-1}$, so as to make the trigonometric function (2) equal to x for all the observed values of t. Thus any variable observed for $2n-1$ equidistant values of t can be represented exactly for these values of t by a series of $2n-1$ harmonic terms. This is the method known as harmonic or Fourier analysis.

E·2. We notice that the possibility of this representation is in no way dependent on the existence of any recognizable physical connexion between x and t. A completely random set of $2n-1$ observations, subjected to harmonic analysis, will give $2n-1$ harmonic terms. Thus if we determine the Fourier coefficients for any definite period, and find them different from zero, the result by itself gives no information that enables us to say that x is, even in part, related to t by any recognizable law.

If $\bar{x}$ be the mean value of x, and if the deviations of x from $\bar{x}$ were quite unrelated to $\cos r\theta$ and $\sin r\theta$, we should still find that the a's and b's did not vanish. This indeed is obvious; for if they did vanish, our theorem shows that x would be the same for all the observed values of t, which is not the case.

Now clearly $\bar{x} = a_0$. Let $x - a_0 = y$. Then

$$\begin{aligned}\sum_{p=0}^{2n-2} y_p^2 &= \sum_{p=0}^{2n-2} (a_1 \cos p\alpha + b_1 \sin p\alpha + \ldots + a_r \cos rp\alpha + b_r \sin rp\alpha + \ldots)^2\\
&= (2n-1)\{\tfrac{1}{2}(a_1^2+b_1^2) + \ldots + \tfrac{1}{2}(a_r^2+b_r^2) + \ldots\}\\
&\quad + \tfrac{1}{2}\sum_p (a_1^2 \cos 2p\alpha + \ldots + a_r^2 \cos 2rp\alpha + \ldots)\\
&\quad - \tfrac{1}{2}\sum_p (b_1^2 \cos 2p\alpha + \ldots + b_r^2 \cos 2rp\alpha + \ldots)\\
&\quad + \text{other trigonometric terms.}\end{aligned}$$

All the trigonometric terms give zero on addition. Thus

$$\sum_r (a_r^2 + b_r^2) = \frac{2}{2n-1}\,\Sigma y_p^2.$$

If then σ denote the standard deviation of x, or of y, so that

$$(2n-1)\,\sigma^2 = \Sigma y_p^2,$$

we have
$$\sum_r (a_r^2 + b_r^2) = 2\sigma^2.$$

Now $a_r^2 + b_r^2$ is the square of the whole amplitude of the term with period $(2n-1)\,T/r$. The probable amplitude is the same for every term. Thus the mean square amplitude of all the terms is $\sigma\sqrt{2/(2n-1)}$.

If then a harmonic analysis showed an amplitude $\sigma\sqrt{2/(2n-1)}$ for every term, we should have no reason for regarding any term as representing anything but the unexplained residuals present in any comparison of theory with observation. We can only assert a term to be 'real' if its amplitude very substantially exceeds this standard.

On the other hand, if the observations exactly fitted a harmonic curve of amplitude ρ, we should have

$$(2n-1)\,\sigma^2 = \Sigma\,\rho^2 \sin^2 r\alpha = \frac{2n-1}{2}\,\rho^2.$$

Thus complete accordance with the harmonic law will exist if the calculated amplitude is $\sigma\sqrt{2}$.

Thus our test of a term is largely determined by the magnitude of its amplitude. If this is equal to $\sigma\sqrt{2/(2n-1)}$ we can make no inference as to whether the term has any physical significance; if it is equal to $\sigma\sqrt{2}$, we can be certain that it expresses the true relation between x and t. All intermediate grades of probability are possible. As a rough working rule, we may say that, if n is large, a term is probably genuine if its amplitude is as great as $3\sigma\sqrt{2/(2n-1)}$, and almost certainly genuine if it reaches $5\sigma\sqrt{2/(2n-1)}$*.

E·21. The periods of all the harmonic terms determined in E·1 are submultiples of $(2n-1)\,T$. Suppose now that x was really a harmonic function of $\lambda\theta$, where λ is between m and $m+1$, m being an integer, and let us examine the values of the harmonic coefficients that would be found. For this purpose it is sufficient to replace the finite summations by integrals. If we are given

$$x = a\cos\lambda\theta + b\sin\lambda\theta,$$

we shall have

$$\pi a_r = \int_0^{2\pi} x\cos r\theta\,d\theta,$$

$$\pi b_r = \int_0^{2\pi} x\sin r\theta\,d\theta,$$

whence

$$2\pi a_r = a\left[\frac{1}{\lambda-r}\sin 2\pi(\lambda-r) + \frac{1}{\lambda+r}\sin 2\pi(\lambda+r)\right]$$
$$+ b\left[\frac{1}{\lambda-r}\{1-\cos 2\pi(\lambda-r)\} + \frac{1}{\lambda+r}\{1-\cos 2\pi(\lambda+r)\}\right],$$

$$2\pi b_r = -a\left[\frac{1}{\lambda-r}\{1-\cos 2\pi(\lambda-r)\} - \frac{1}{\lambda+r}\{1-\cos 2\pi(\lambda+r)\}\right]$$
$$+ b\left[\frac{1}{\lambda-r}\sin 2\pi(\lambda-r) - \frac{1}{\lambda+r}\sin 2\pi(\lambda+r)\right].$$

The terms depending on $2\pi(\lambda+r)$ are evidently at most of order $\frac{1}{r}$, but those involving $2\pi(\lambda-r)$ may be of order unity. Hence if the harmonics

* Cf. Sir A. Schuster, *Proc. Roy. Soc.* **61** A, 1897, 455–465. It is shown that if n events are distributed at random and their frequencies harmonically analysed, the expectancy for the ratio ρ/a_0 is $(\pi/n)^{\frac{1}{2}}$. This may be used instead of $\sigma\sqrt{2/(2n-1)}$ when frequencies are analysed.

considered are of high order, the coefficients of the largest terms may be replaced by

$$\pi a_r = \frac{\sin \pi (\lambda - r)}{\lambda - r} \{a \cos \pi (\lambda - r) + b \sin \pi (\lambda - r)\},$$

$$\pi b_r = \frac{\sin \pi (\lambda - r)}{\lambda - r} \{- a \sin \pi (\lambda - r) + b \cos \pi (\lambda - r)\},$$

and
$$\pi (a_r^2 + b_r^2)^{\frac{1}{2}} = \frac{\sin \pi (\lambda - r)}{\lambda - r} (a^2 + b^2)^{\frac{1}{2}}.$$

If then λ is equal to $r + \frac{1}{2}$, the two largest terms found by the harmonic analysis will be equal in amplitude. For other values, the values of the amplitudes found for the terms in $r\theta$ and $(r + 1)\,\theta$ will suffice to determine $(a^2 + b^2)^{\frac{1}{2}}$ and λ. Thus an extension of the method of harmonic analysis makes it possible to find approximately the amplitude and period of a harmonic variation, even though its period may not be a submultiple of $(2n - 1)\,T$. Thus the results of the harmonic analysis serve as a compendium whence the harmonic variations with periods intermediate between those used in the analysis can be inferred.

E·3. Supposing the existence of a harmonic term to be established by analysis, an interpretation of this term is required. The character of this interpretation is very different in different cases. The simplest type is represented by a forced vibration where the external variation producing the motion is known beforehand to have a definite period. The phenomena of the tides provide an excellent example. The disturbing potentials due to the moon and sun have been expressed as the sum of a number of terms of different periods. Theoretically, each of these should produce variations in the height of the tide and the velocity of the current at any given place, with the period of the disturbing term that produces them; and the total height and velocity are the resultants of those due to the terms separately. Thus a harmonic analysis of the tidal observations at a given station enables each constituent separately to be found, and the results can afterwards be used to predict the tides at the station at any subsequent time. In this form the method presents no methodological difficulty: it is used only to determine the amplitudes and phases of terms that we already know must exist.

The next type in order of difficulty is the determination of the period of a free vibration. We may have previous reason to suspect the existence of a free vibration, and our theory may be so reliable that we can infer its period with considerable accuracy. If this is so, the determination of the amplitude and phase presents no more difficulty than in the case of a forced vibration. But in most geophysical phenomena the free periods are not accurately known beforehand. All we know is that a certain type of oscillation is possible, and perhaps we may know in addition the order of magnitude of its period. Thus to find the period we must carry out a harmonic analysis of the observations to find out what period agrees best with them. The analysis is complicated by the fact that free vibrations as a rule are affected by damping, the amount of the latter being usually unknown beforehand. This is innocuous in a forced vibration, for it does not affect the period, but only the amplitude and phase, which we find directly. But in a free vibration it affects the period and makes the amplitude subject to a steady diminution. If a further disturbance regenerates

the vibration, the new vibration will not necessarily be in the same phase as the old one, and therefore an analysis covering both may give quite incorrect estimates of the amplitude and even the period. This difficulty has already been noted in connexion with the 14-monthly variation of latitude. Again, we may have no strong theoretical reason to believe that the free vibration we are seeking actually exists, for it may have been completely damped out, even if, indeed, it ever started. In such a case we cannot exclude the possibility that an empirical term with a period agreeing only roughly with our preliminary estimate may be due to some completely different cause. Again, it may happen that several periods, all of the correct order of magnitude, are disclosed by the analysis. If so, it will be difficult to say definitely which of them represents the predicted free vibration.

The next degree of difficulty occurs when we have no previous reason whatever to believe in the existence of a variation with a period comparable with that found. The 11-year period of sunspots is an example. Nobody knows why the number of spots on the sun should vary; we only know that it does. There is an equal period in the frequency of magnetic storms, and it is natural to infer from the agreement between the periods that the two phenomena are connected by a physical law. Now in such cases as these the actual variation is never simple harmonic; other periods and an irregular part are always present. If there is a physical connexion, we should expect it to hold, say, between the mean sunspot number for the month and the whole number of magnetic storms in the month, and not merely between the 11-yearly parts of the two variations. Thus it is natural to inquire why the method of harmonic analysis should in such cases be preferred to the method of correlation.

The answer appears to be as follows. The harmonic analysis enables us to find sets of harmonic terms that will represent any variation exactly. If then we analyse the two phenomena as in E·1, and consider the terms of any two periods P and Q, the average values of the products of the terms of period P in the sunspot number into the terms of period Q in the number of magnetic storms are zero. Thus the combinations of terms of different periods contribute nothing to the correlation coefficient. On the other hand, if we take together the terms of period P in both, their products will as a rule not vanish when averaged over a long period, and they will therefore contribute to the correlation coefficient. Thus a harmonic analysis tells us all that the correlation coefficient can tell us; but it also tells us what part of the correlation is due to every periodic term separately. Further, it gives the ratios of the amplitudes and the lags of the separate terms. A knowledge of these quantities, and of the way they vary with the period, must be helpful in elucidating the nature of the connexion between the phenomena considered.

The harmonic analysis of a single series of observations, whose components are not to be compared with those of any other series, does not appear to be necessarily of any scientific value. The case of a predicted free vibration is hardly a case in point, since the results of the analysis are to be compared with theoretical considerations based on other data. In any investigation of a series of observations four cases may arise. First, it may happen that the existence of a harmonic term is predicted by theoretical considerations, and that such a term is found to give a good representation of the observations. In this case, typified by the tides, harmonic analysis

achieves its greatest possible success. Second, it may be found that the observations are well represented by a harmonic term or by a few harmonic terms, but we may have no theoretical explanation of the existence of such terms. In this case the analysis supplies us with valuable data for future research, but not with knowledge relevant to any laws. Third, theory may predict the existence of a harmonic term, but on examination of the data it may be found that the harmonic term, though its existence may be well established, actually accounts for only a small fraction of the whole range of variation. This is specially liable to happen when long series of observations are analysed. The lunar tide in the atmosphere, investigated by S. Chapman, affords a striking example. Such a result constitutes a verification of a theory, but it is evidently useless for prediction of individual observations. The fourth possibility is that we have no reason to expect the existence of a periodicity, and that on investigation we find that the periodicity does not represent more than a small fraction of the observed range of variation. In such a case we only find by analysis that what we had no reason to expect to happen does not in fact happen, which does not appear to be a result of any scientific interest. Many examples of such 'periodicities' have been published.

E·4. For our present purpose the most interesting periodicities are those in the frequencies of earthquakes. A great deal of work has been done on these periodicities, but so far only two appear to be well established. These are the solar diurnal and annual variations* found by C. Davison. It is found that, on the whole, earthquakes are more frequent in winter than in summer, and more frequent by night than by day.

It appears that these variations are of the nature of "last straw phenomena." Under the growing stresses in the earth's crust, fracture is bound to occur sooner or later. A small periodic variation in stress, however, may determine the time when the fracture actually takes place, provided the range of the variation is comparable with the amount of the steady increase during a period. If, for instance, the stress-differences in a region are greater, on the whole, in winter than in summer, they will be decreasing from winter to summer even when the steady increase is taken into account. Therefore there will be no earthquakes during this half-year; all will occur in the half of the year from summer to winter. Davison, largely following Omori, is inclined to attribute the annual period to the annual variation of atmospheric pressure, and the diurnal period to the diurnal variation of atmospheric pressure.

That these variations are not tidal in origin may be seen from the fact that the corresponding lunar terms are very much smaller, their amplitudes being little more than would be expected if the total variations were random. They have been the subject of investigation by R. D. Oldham†, and before him by C. G. Knott. The smallness of these lunar variations is, however, as interesting as their verification would have been. The extensional strain in the earth's crust since solidification has been seen to be of the order of 1 per cent., which has accumulated in a time of the order of 10^9 years. Thus the increase of extension in a month is of the order of 10^{-12}. The extension produced by the lunar bodily tide, how-

* C. Davison, *A Manual of Seismology*, Camb. Univ. Press, 1921, 82–198. References to papers are given.

† *Q. Journ. Geol. Soc.* **74**, 1918–19, 99–104; **77**, 1921, 1–3; **78**, 1922, lv–lxii.

ever, must be of the order of 10^{-6}. It therefore becomes of interest to inquire why the lunar periodicity does not so dominate the occurrence of earthquakes as to make all the earthquakes in any region occur in only one half of the lunar day. The difficulty has been noticed by Oldham, who suggests that, while the gradual increase of stress produces set ultimately, the set is at first gradual and does not give a definite earthquake. When set has begun, however, it proceeds rapidly, and the small lunar effect is unable to influence appreciably the moment when the flow merges into fracture.

Another periodicity of much interest has been noticed by Prof. H. H. Turner*. The aftershocks of earthquakes occur at intervals of multiples of 21 minutes. In a further development Jeans† points out that the intervals are equally well represented by multiples of 126 minutes and 222 minutes, and combinations of the two. As a result of an investigation *ab initio* of the theory of earthquake wave propagation, he finds that three types of surface waves are possible. When the wave length is sufficiently small, the first type travels with the velocity of P waves in the surface rocks, the second with the velocity of S waves, and the third with the velocity of Rayleigh waves in those rocks. The second type are evidently Love waves and the third Rayleigh waves, of very short wave length in each case. Jeans suggests that when the first of these waves has travelled right round the earth to the starting-point, it starts a new earthquake, and that the process may be repeated. The second wave may also play a part.

The theory is suggestive, but is open to some objections. If the wave length is very short, it appears that the wave should be confined to a very small depth, and it is not easy to see how it can produce any shock at the depth where the aftershocks originate. Again, the times taken by the waves to travel round the earth are inferred from the velocities found by Dr Wrinch and myself from the Oppau explosion. These velocities definitely refer to the granitic layer of the continents. If the surface rocks of the ocean are basaltic, as is probable, the velocities in them will be different, and the quantitative agreement will break down. Further, the times of circuit will be different in different azimuths, and thus the waves will not be focused at the epicentre. These waves, again, have not been recorded at great distances, and it is difficult to see why the ordinary long waves, which are much larger and are faster than Jeans's second type, do not produce more aftershocks. Nevertheless Jeans appears to have discovered a genuine phenomenon, which will probably remain among the data of seismology even if his explanation should fail to meet the objections offered.

E·5. A further set of periodicities has been discovered by Prof. Turner‡ in the frequency of Chinese earthquakes, the heights of the Nile floods, the growth of Californian Sequoias, and the moon's longitude. Each of these phenomena shows a periodicity in a period of 250 to 300 years, those for the Californian trees being the best analysed because they provide the longest record. The records are such as to show such a period fairly definitely in all, and the phases agree as closely as the analysis can detect. No satisfactory explanation of the relations has been suggested.

* *M N.R.A.S. Geophys. Suppt.* 1, 1923, 31–50. † *Proc. Roy. Soc.* **102** A, 1923, 554–574.

‡ H. H. Turner, *M.N.R.A.S.* **79**, 1919, 531–9; **80**, 1920, 617–9, 793–808. A. E. Douglass, *Climatic Cycles and Tree Growth*, Carnegie Institution Publ. No. 289, 1919, pp. 127.

INDEX

CAMBRIDGE: PRINTED BY W. LEWIS AT THE UNIVERSITY PRESS